MULTIPLE RECURRENCE AND CONVERGENCE FOR HARDY SEQUENCES OF POLYNOMIAL GROWTH

NIKOS FRANTZIKINAKIS

Abstract. We study the limiting behavior of multiple ergodic averages involving sequences of integers that satisfy some regularity conditions and have polynomial growth. We show that for "typical" choices of Hardy field functions $a(t)$ with polynomial growth, the averages $\frac{1}{N}\sum_{n=1}^N f_1(T^{[a(n)]}x)\cdot\ldots\cdot f_\ell(T^{\ell[a(n)]}x)$ converge in the mean and we determine their limit. For example, this is the case if $a(t)=t^{3/2}, t\log t$, or $t^2+(\log t)^2$. Furthermore, if $\{a_1(t),\ldots,a_\ell(t)\}$ is a "typical" family of logarithmico-exponential functions of polynomial growth, then for every ergodic system, the averages $\frac{1}{N}\sum_{n=1}^N f_1(T^{[a_1(n)]}x)\cdot\ldots\cdot f_\ell(T^{[a_\ell(n)]}x)$ converge in the mean to the product of the integrals of the corresponding functions. For example, this is the case if the functions $a_i(t)$ are given by different positive fractional powers of t. We deduce several results in combinatorics. We show that if $a(t)$ is a non-polynomial Hardy field function with polynomial growth, then every set of integers with positive upper density contains arithmetic progressions of the form $\{m, m+[a(n)],\ldots,m+\ell[a(n)]\}$. Under suitable assumptions we get a related result concerning patterns of the form $\{m,m+[a_1(n)],\ldots,m+[a_\ell(n)]\}$.

April 27, 2010

Contents

1. Introduction

In recent years there has been a lot of activity in studying the limiting behavior in $L^2(\mu)$ (as $N\to\infty$) of multiple ergodic averages of the form

$$\frac{1}{N}\sum_{n=1}^N f_1(T^{a_1(n)}x)\cdot\ldots\cdot f_\ell(T^{a_\ell(n)}x) \tag{1}$$

for various choices of sequences of integers $a_1(n),\ldots,a_\ell(n)$, where T is an invertible measure preserving transformation acting on a probability space $(X,\mathcal{X},\mu)$, and $f_1,\ldots,f_\ell$ are bounded

2000 *Mathematics Subject Classification.* Primary: 37A45; Secondary: 28D05, 05D10, 11B25.
Key words and phrases. Hardy field, multiple recurrence, ergodic averages, arithmetic progressions.
The author was partially supported by NSF grant DMS-0701027 and Marie Curie International Reintegration Grant 248008.

measurable functions. This study was initiated in [18], where Furstenberg studied the averages (1) when $a_1(n) = n, a_2(n) = 2n, \ldots, a_\ell(n) = \ell n$, in a depth that was sufficient to give a new proof of Szemerédi's theorem on arithmetic progressions ([32]). Later on, Bergelson and Leibman in [9] extended Furstenberg's method to cover the case where the sequences $a_1(n), \ldots, a_\ell(n)$ are integer polynomials with zero constant term, and established a polynomial extension of Szemerédi's theorem.

For $\ell = 1$ the limiting behavior in $L^2(\mu)$ of the averages (1) can be reduced (using the spectral theorem for unitary operators) to the study of certain exponential sums, and therefore is in a sense well understood. For $\ell \geq 2$, even in the simplest cases, convergence of the averages (1) and identification of the limit turned out to be a very resistent problem. Nevertheless, we now have several different proofs of convergence in the case where the sequences are linear ([26], [38], [33], [35],[1], [25]), and the case where all the sequences are integer polynomials was treated in [27] and [29]. Furthermore, a rather explicit formula for the limit of the averages (1) can be given in the linear case (combining results from [26] and [37]), and for some special collections of integer polynomial sequences ([16], [13], [30]).

The purpose of this article is to carry out a detailed study of the limiting behavior of the averages (1) for a large class of sequences of integers $a_1(n), \ldots, a_\ell(n)$ that have polynomial growth (meaning $a_i(t)/t^k \to 0$ for some $k \in \mathbb{N}$) but are not necessarily defined by integer polynomials.[1] For example, we shall show that for every positive $c \in \mathbb{R} \setminus \mathbb{Z}$, and measure preserving transformation T, the averages

$$\frac{1}{N}\sum_{n=1}^{N} f_1(T^{[n^c]}x) \cdot f_2(T^{2[n^c]}x) \cdot \ldots \cdot f_\ell(T^{\ell[n^c]}x) \tag{2}$$

converge in $L^2(\mu)$, and their limit is equal to the limit of the "Furstenberg averages"

$$\frac{1}{N}\sum_{n=1}^{N} f_1(T^{n}x) \cdot f_2(T^{2n}x) \cdot \ldots \cdot f_\ell(T^{\ell n}x). \tag{3}$$

More generally, the role of the sequence $[n^c]$ in (2) can play any sequence $[a(n)]$ where $a(t)$ is a function that belongs to some Hardy field, has polynomial growth, and stays logarithmically away from constant multiples of integer polynomials (see Theorem 2.2). For instance, any of the following sequences will work (below, k is an arbitrary positive integer)

$$[n\log n],\ [n^3/\log n],\ [n^2+n\log n],\ [n^2+\sqrt{3}\,n],\ [n^2+(\log n)^2],\ [(\log(n!))^k],\ [(\mathrm{Li}(n))^k]. \tag{4}$$

We also give explicit necessary and sufficient conditions for mean convergence of the averages (2) when the sequence $[n^c]$ is replaced with the sequence $[a(n)]$ where $a(t)$ is any function that belongs to some Hardy field and has polynomial growth.

With the help of the previous convergence results we derive a refinement of Szemerédi's theorem on arithmetic progressions. We show that if $a(t)$ is a function that belongs to some Hardy field, has polynomial growth, and is not equal to a constant multiple of an integer polynomial (modulo a function that converges to a constant), then for every $\ell \in \mathbb{N}$, every set of integers with positive upper density[2] contains arithmetic progressions of the form

$$\{m, m+[a(n)], \ldots, m+\ell[a(n)]\}$$

[1] The case where the transformation T is a nilrotation was treated in the companion paper [14] and is an essential component of the present paper.

[2] A set of integers Λ has *positive upper density* if $\bar{d}(\Lambda) = \limsup_{N\to\infty} |\Lambda \cap \{-N, \ldots, N\}|/(2N+1) > 0$.

(see Theorem 2.5). Therefore, one can use any of the sequences in (4) in place of $[a(n)]$ and also sequences like $[n^2+\log n]$ or $[\sqrt{2}n^2+\log\log n]$ (these sequences are bad for mean convergence).

Furthermore, we study the averages (1) for sequences that are not necessarily in arithmetic progression. We show that if $c_1,\ldots,c_\ell\in\mathbb{R}\setminus\mathbb{Z}$ are positive and distinct, then for every ergodic transformation T we have

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N} f_1(T^{[n^{c_1}]}x)\cdot\ldots\cdot f_\ell(T^{[n^{c_\ell}]}x)=\int f_1\; d\mu\cdot\ldots\cdot\int f_\ell\; d\mu \tag{5}$$

where the convergence takes place in $L^2(\mu)$. More generally, one can replace the sequences $[n^{c_1}],\ldots,[n^{c_\ell}]$ in (5) with sequences $[a_1(n)],\ldots,[a_\ell(n)]$ where the functions $a_1(t),\ldots,a_\ell(t)$ are logarithmico-exponential and satisfy some appropriate growth conditions (see Theorem 2.6). This enables us to establish a conjecture of Bergelson and Håland-Knutson ([7], Conjecture 8.2). We deduce that if $c_1,\ldots,c_\ell\in\mathbb{R}\setminus\mathbb{Z}$ are positive, then every set of integers with positive upper density contains patterns of the form

$$\{m,m+[n^{c_1}],\ldots,m+[n^{c_\ell}]\}.$$

In the next section we give a more precise formulation of our main results and also define some of the concepts used throughout the paper.

2. Main results

We first introduce some basic terminology needed to state our main results. The reader will find more information related to the notions involved in the background section.

All along the article we will use the term *measure preserving system*, or the word *system*, to designate a quadruple $(X,\mathcal{X},\mu,T)$, where $(X,\mathcal{X},\mu)$ is a Lebesgue probability space, and $T\colon X\to X$ is an *invertible* measurable map such that $\mu(T^{-1}A)=\mu(A)$ for every $A\in\mathcal{X}$.

Let B be the collection of equivalence classes of real valued functions defined on some half-line (c,∞), where we identify two functions if they agree eventually.[3] A *Hardy field* is a subfield of the ring $(B,+,\cdot)$ that is closed under differentiation. With $\mathcal{H}$ we denote the *union of all Hardy fields*. If $a\in\mathcal{H}$ is defined in $[1,\infty)$ (one can always choose such a representative of $a(t)$) we call the sequence $([a(n)])_{n\in\mathbb{N}}$ a *Hardy sequence*. Working within the class $\mathcal{H}$ eliminates several technicalities that would otherwise obscure the transparency of our results and the main ideas of their proofs. Furthermore, $\mathcal{H}$ is a rich enough class to enable one to deal, for example, with all the sequences considered in (4).

An explicit example of a Hardy field to keep in mind is the set $\mathcal{LE}$ that consists of all *logarithmico-exponential functions* ([23], [24]), meaning all functions defined on some half-line (c,∞) using a finite combination of the symbols $+,-,\times,:,\log,\exp$, operating on the real variable t and on real constants. For example, all rational functions and the functions $t^{\sqrt{2}}$, $t\log t$, $e^{\sqrt{\log\log t}}/\log(t^2+1)$ belong in $\mathcal{LE}$.

The set $\mathcal{H}$ is much more extensive than the set $\mathcal{LE}$, for example, one can show that it contains all antiderivatives of elements of $\mathcal{LE}$, the Riemann zeta function ζ, and the Euler Gamma function Γ. Let us stress though that $\mathcal{H}$ does not contain functions that oscillate like $\sin t$ or $t\sin t$, or functions that have a derivative that oscillates, like $t^{100}+\sin t$.

[3]The equivalence classes just defined are often called *germs of functions*. We are going to use the word function when we refer to elements of B instead, with the understanding that all the operations defined and statements made for elements of B are considered only for sufficiently large values of $t\in\mathbb{R}$.

The most important property of elements of $\mathcal{H}$ that will be used throughout this article is that we can relate their growth rates with the growth rates of their derivatives.

To simplify our exposition we introduce some notation. If $a(t), b(t)$ are real valued functions defined on some half-line (u, ∞) we write $a(t) \prec b(t)$ if $a(t)/b(t) \to 0$ as $t \to \infty$. (For example, $1 \prec \log t \prec t^{\varepsilon}$ for every $\varepsilon > 0$.) We write $a(t) \ll b(t)$ if there exists $C \in \mathbb{R}$ such that $|a(t)| \leq C|b(t)|$ for all large enough $t \in \mathbb{R}$. We say that a function $a(t)$ has *polynomial growth* if $a(t) \ll t^k$ for some $k \in \mathbb{N}$.

2.1. **Arithmetic progressions.** We are going to give a rather exhaustive collection of results that deal with multiple convergence and recurrence properties of Hardy sequences $([a(n)])_{n\in\mathbb{N}}$ of polynomial growth.

2.1.1. *Convergence.* Let $a(t)$ be a real valued function. We say that the sequence of integers $([a(n)])_{n\in\mathbb{N}}$ is *good for multiple convergence* if for every $\ell \in \mathbb{N}$, system $(X, \mathcal{X}, \mu, T)$, and functions $f_1, f_2, \dots, f_\ell \in L^\infty(\mu)$, the averages

$$\frac{1}{N}\sum_{n=1}^{N} f_1(T^{[a(n)]}x) \cdot f_2(T^{2[a(n)]}x) \cdot \ldots \cdot f_\ell(T^{\ell[a(n)]}x) \tag{6}$$

converge in $L^2(\mu)$ as $N \to \infty$. As mentioned in the introduction, any polynomial with integer coefficients is an example of such a sequence.

The next result gives an extensive list of new examples of sequences that are good for multiple convergence. In fact it shows that it is a rather rare occurrence for a Hardy sequence with polynomial growth to be bad for multiple convergence.

Theorem 2.1. *Let $a \in \mathcal{H}$ have polynomial growth.*

Then the sequence $([a(n)])_{n\in\mathbb{N}}$ is good for multiple convergence if and only if one of the following conditions holds:

- $|a(t) - cp(t)| \succ \log t$ *for every* $c \in \mathbb{R}$ *and every* $p \in \mathbb{Z}[t]$; *or*
- $a(t) - cp(t) \to d$ *for some* $c, d \in \mathbb{R}$ *and some* $p \in \mathbb{Z}[t]$; *or*
- $|a(t) - t/m| \ll \log t$ *for some* $m \in \mathbb{Z}$.

Remarks. • It follows that the sequences in (4) and the sequences $[\sqrt{5}n^2]$, $[n/2 + \log n]$ are good for multiple convergence. The sequences $[\sqrt{5}n^2 + \log n]$, $[2n + \log n]$ are bad for multiple convergence.

• The same necessary and sufficient conditions for convergence of "single" ergodic averages $\frac{1}{N}\sum_{n=1}^{N} f(T^{[a(n)]}x)$ where previously established in [12].

• If $a(t)$ is a real valued polynomial, then our argument shows that the averages (6) converge in $L^2(\mu)$ even if one replaces the limit $\lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N}$ with the limit $\lim_{N-M\to\infty} \frac{1}{N-M}\sum_{n=M+1}^{N}$. On the other hand, if $a \in \mathcal{H}$ satisfies $t^{k-1} \prec a(t) \prec t^k$ for some $k \in \mathbb{N}$, then one can show that the sequence $([a(n)])_{n\in\mathbb{N}}$ takes odd (respectively even) values in arbitrarily long intervals; as a result the limit $\lim_{N-M\to\infty} \frac{1}{N-M}\sum_{n=M+1}^{N} T^{[a(n)]}f$ does not exist in general.

Notice that the first condition of Theorem 2.1 is satisfied by the "typical" function in $\mathcal{H}$ with polynomial growth. The next result allows us to identify the limit of the averages (6) for such "typical" functions:

Theorem 2.2. *Let $a \in \mathcal{H}$ have polynomial growth and satisfy $|a(t) - cp(t)| \succ \log t$ for every $c \in \mathbb{R}$ and every $p \in \mathbb{Z}[t]$.*

Then for every $\ell \in \mathbb{N}$, system $(X, \mathcal{X}, \mu, T)$, and functions $f_1, \ldots, f_\ell \in L^\infty(\mu)$, we have

$$(7)\qquad \lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} T^{[a(n)]}f_1 \cdot T^{2[a(n)]}f_2 \cdot \ldots \cdot T^{\ell[a(n)]}f_\ell = \lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} T^{n}f_1 \cdot T^{2n}f_2 \cdot \ldots \cdot T^{\ell n}f_\ell$$

where the limit is taken in $L^2(\mu)$.

Remarks. • Examples of Hardy sequences for which this result applies are those given in (4).

• A rather explicit formula for the limit in (7) can be given by combining results in [26] (see also [38]) and [37].

• If $a(t) = cp(t) + d$ for some $c \in \mathbb{R}$ and $p \in \mathbb{Z}[t]$, then (7) typically fails. One can see this by considering appropriate rotations on the circle and taking $a(t) = 2t, t^2$, or $\sqrt{2}t$.

2.1.2. *Recurrence.* Let $a(t)$ be a real valued function. We say that the sequence of integers $([a(n)])_{n\in\mathbb{N}}$ is *good for multiple recurrence* if for every $\ell \in \mathbb{N}$, system $(X, \mathcal{X}, \mu, T)$, and set $A \in \mathcal{X}$ with $\mu(A) > 0$ we have

$$(8)\qquad \mu(A \cap T^{-[a(n)]}A \cap T^{-2[a(n)]}A \cap \cdots \cap T^{-\ell[a(n)]}A) > 0$$

for some $n \in \mathbb{N}$ such that $[a(n)] \neq 0$. One can check that if the sequence $([a(n)])_{n\in\mathbb{N}}$ is good for multiple recurrence, then (8) is satisfied for infinitely many $n \in \mathbb{N}$.

Let us discuss briefly the recurrence properties of sequences defined using polynomials with real coefficients. If $q \in \mathbb{R}[t]$ is non-constant and has zero constant term, then the sequence $q(n)$ is good for multiple recurrence (this follows from [9] and a trick used in [6]). If $q \in \mathbb{Z}[t]$ does not have zero constant term, then the sequence $q(n)$ is good for multiple recurrence if and only if the range of the polynomial contains multiples of every positive integer ([13]). More generally, if $q \in \mathbb{R}[t]$, then $[q(n)]$ is good for multiple recurrence unless $q(t)$ has the form $q(t) = cp(t) + d$ for some $p \in \mathbb{Z}[t]$ and $c, d \in \mathbb{R}$ (one way to see this is to use Theorem 2.3 below). In this last case deciding whether the sequence $[q(n)]$ is good for multiple recurrence is more delicate and depends on intrinsic properties of the polynomial q. For example, one can show that the sequences $[\sqrt{5}n+1]$ and $[\sqrt{5}n+3]$ are good for multiple recurrence, but the sequence $[\sqrt{5}n+2]$ is bad for multiple recurrence.[4]

Our next result shows that if one avoids polynomial sequences, then every Hardy sequence of polynomial growth is good for multiple recurrence:

Theorem 2.3. *Let $a \in \mathcal{H}$ have polynomial growth and suppose that $a(t) - cp(t) \to \infty$ for every $c \in \mathbb{R}$ and $p \in \mathbb{Z}[t]$.*

Then the sequence $([a(n)])_{n\in\mathbb{N}}$ is good for multiple recurrence.

Remarks. • Examples of Hardy sequences for which this result applies are given in (4). It also applies to the sequences $[\sqrt{5}n + \log n]$ and $[n^2 + \log\log n]$.

• Theorem 2.3 was previously established in [17] under a somewhat more restrictive assumption (namely $t^{k-1} \prec a(t) \prec t^k$ for some $k \in \mathbb{N}$). Furthermore, the single recurrence case was previously established by Boshernitzan (unpublished), and subsequently in [17].

• Let R be the set of those $n \in \mathbb{N}$ for which (8) holds. Combining the multiple recurrence result of Furstenberg ([18]) and Theorem 2.2, one sees that if $a(t) - cp(t) \succ \log t$ for every

[4] The sequence $[\sqrt{5}n+2]$ is bad for recurrence because $\|[\sqrt{5}n+2]/\sqrt{5}\| \geq 1/10$ for every $n \in \mathbb{N}$, where $\|\cdot\|$ denotes the distance to the closest integer. It can be shown ([6]) that the sequence $[an+b]$, $a, b \in \mathbb{R}$, is good for single recurrence (meaning (8) holds for $\ell = 1$) if and only if there exists an integer k such that $ak + b \in [0, 1]$ (this is equivalent to $\{b/a\} \leq 1/a$). For other sequences of the form $[ap(n) + n]$, like $[an^2 + b]$, necessary and sufficient conditions seem to be more complicated.

$c \in \mathbb{R}$ and $p \in \mathbb{Z}[t]$, then the set R has positive lower density. Unlike the case where $a(t)$ is polynomial, if $a \in \mathcal{H}$ satisfies $t^{k-1} \prec a(t) \prec t^k$ for some $k \in \mathbb{N}$, then one can show that the sequence $[a(n)]$ takes odd values in arbitrarily long intervals, and as a result for some systems the set R has unbounded gaps.

2.1.3. *Characteristic factors.* Let $(X, \mathcal{X}, \mu, T)$ be a system. A factor $\mathcal{C}$ is called a *characteristic factor*, or *characteristic*, for the family of integer sequences $\{a_1(n), \ldots, a_\ell(n)\}$, if whenever one of the functions $f_1, \ldots, f_\ell \in L^\infty(\mu)$ is orthogonal to $\mathcal{C}$, the averages

$$\frac{1}{N}\sum_{n=1}^{N} f_1(T^{a_1(n)}x) \cdot \ldots \cdot f_\ell(T^{a_\ell(n)}x) \tag{9}$$

converge to 0 in $L^2(\mu)$ as $N \to \infty$.

It follows that if $\mathcal{C}$ is as above, then the limiting behavior of the averages (9) remains unchanged if one projects each function to the factor $\mathcal{C}$, meaning that the difference of the two averages converges to 0 in $L^2(\mu)$ as $N \to \infty$.

It is known that the nilfactor $\mathcal{Z}$ of a system (defined in Section 3.3) is characteristic for every family $\{p(n), 2p(n), \ldots, \ell p(n)\}$ whenever p is an integer polynomial ([27]). We extend this result by showing:

Theorem 2.4. *Suppose that $a \in \mathcal{H}$ has polynomial growth and satisfies $a(t) \succ \log t$.*

Then for every system and $\ell \in \mathbb{N}$, the nilfactor $\mathcal{Z}$ of the system is characteristic for the family $\{[a(n)], 2[a(n)], \ldots, \ell[a(n)]\}$, for every $\ell \in \mathbb{N}$.

Remarks. • If $a(t) \ll \log t$, then the result fails even for $\ell = 1$, the reason being that the sequence $[a(n)]$ remains constant on some sub-interval of $[1, N]$ that has length proportional to N as $N \to \infty$. Therefore, if the transformation T is weakly mixing but not strongly mixing, the function f has zero integral and satisfies $\int f \cdot T^n f \, d\mu \not\to 0$, then $f \perp \mathcal{Z}$ but $\frac{1}{N}\sum_{n=1}^{N} T^{[a(n)]} f \not\to 0$.

• A related result was proved in [7] for weakly mixing systems assuming that the function $a(t)$ is tempered (for $a \in \mathcal{H}$ this is equivalent to $t^{k-1}\log t \prec a(t) \prec t^k$ for some $k \in \mathbb{N}$). Since the method used in [7] does not work for functions like $t^k \log t$, we will use a different approach to prove Theorem 2.4.

2.1.4. *Combinatorics.* Using the previous multiple recurrence result we derive a refinement of Szemerédi's Theorem on arithmetic progressions. We will use the following correspondence principle of Furstenberg (the formulation given is from [4]):

Furstenberg Correspondence Principle. *Let Λ be a set of integers.*

Then there exist a system $(X, \mathcal{B}, \mu, T)$ and a set $A \in \mathcal{X}$, with $\mu(A) = \bar{d}(\Lambda)$, and such that

$$\bar{d}(\Lambda \cap (\Lambda - n_1) \cap \ldots \cap (\Lambda - n_\ell)) \geq \mu(A \cap T^{-n_1}A \cap \cdots \cap T^{-n_\ell}A) \tag{10}$$

for every $n_1, \ldots, n_\ell \in \mathbb{Z}$ and $\ell \in \mathbb{N}$.

Using the previous principle and Theorem 2.3 we immediately deduce the following:

Theorem 2.5. *Let $a \in \mathcal{H}$ have polynomial growth and suppose that $a(t) - cp(t) \to \infty$ for every $c \in \mathbb{R}$ and every $p \in \mathbb{Z}[t]$.*

Then for every $\ell \in \mathbb{N}$, every $\Lambda \subset \mathbb{Z}$ with $\bar{d}(\Lambda) > 0$ contains arithmetic progressions of the form

$$\{m, m + [a(n)], m + 2[a(n)], \ldots, m + \ell[a(n)]\} \tag{11}$$

for some $m \in \mathbb{Z}$ and $n \in \mathbb{N}$ with $[a(n)] \neq 0$.

2.1.5. *More general classes of functions.* We make some remarks about the extend of the functions our methods cover that do not necessarily belong to some Hardy field.

The conclusions of Theorems 2.2, 2.3, 2.4, and 2.5 hold if for some $k \in \mathbb{N}$ the function $a \in C^{k+1}(\mathbb{R}_+)$ satisfies

$$|a^{(k+1)}(t)| \text{ decreases to zero}, \quad 1/t^k \prec a^{(k)}(t) \prec 1, \quad \text{and} \quad (a^{(k+1)}(t))^k \prec (a^{(k)}(t))^{k+1}.$$

(If $a \in \mathcal{H}$, these three conditions are equivalent to "$a(t)$ has polynomial growth and $|a(t) - p(t)| \succ \log t$ for every $p \in \mathbb{R}[t]$".) One can see this by repeating verbatim the proofs given in this article and in [14]. The reader is advised to think of the second condition as the most important one and the other two as technical necessities (for functions in $\mathcal{H}$ the second condition implies the other two).

As for Theorem 2.1, unless one works within a "regular" class of functions like $\mathcal{H}$, it seems impossible to get explicit necessary and sufficient conditions.

2.2. **Several sequences.** We are going to give results related to multiple convergence and recurrence properties involving several sequences of polynomial growth. For practical reasons (mainly expository) we are going to restrict ourselves to the case where all the functions involved are logarithmico-exponential. More technically involved arguments should enable one to extend the results mentioned below to the case where all the functions belong to the same Hardy field.

Let us also remark that the results we give below are certainly less exhaustive than the results of Section 2.1. We are able to handle a case that includes all functions given by fractional powers of t and is general enough to cover a conjecture of Bergelson and Håland. The expected "optimal" results involving several sequences are stated in Problems 2, 3, and 4 of Section 2.3.

2.2.1. *Convergence.* To simplify our statements we introduce the following class of "good" (for our purposes) functions:

$$\mathcal{G} = \{a \in C(\mathbb{R}_+)\colon t^{k+\varepsilon} \prec a(t) \prec t^{k+1} \text{ for some integer } k \geq 0 \text{ and some } \varepsilon > 0\}. \tag{12}$$

Equivalently, a function $a \in \mathcal{H}$ with polynomial growth belongs in $\mathcal{G}$ *unless* for some integer $k \geq 0$ we have $t^k \prec a(t) \prec t^{k+\varepsilon}$ for every $\varepsilon > 0$. For example, if $c \geq 0$, then $t^c \in \mathcal{G}$ if and only if c is not an integer. The reader is advised to think of functions in $\mathcal{G}$ as having "fractional-power growth rate".

The next result (in fact its corollary Theorem 2.8) verifies a conjecture of Bergelson and Håland-Knutson ([7], Conjecture 8.2).[5] It shows that for "typical" logarithmico-exponential functions of polynomial growth the limit of the averages (1) exists and for ergodic systems it is constant. We say that the functions $a_1(t), \dots, a_\ell(t)$ have *different growth rates* if the quotient of any two of these functions converges to $\pm\infty$ or to 0.

Theorem 2.6. *Suppose that the functions $a_1, \dots, a_\ell \in \mathcal{LE} \cap \mathcal{G}$ have different growth rates.*

Then for every ergodic system $(X, \mathcal{B}, \mu, T)$ and $f_1, \dots, f_\ell \in L^\infty(\mu)$ we have

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} f_1(T^{[a_1(n)]}x) \cdot \ldots \cdot f_\ell(T^{[a_\ell(n)]}x) = \int f_1 \, d\mu \cdot \ldots \cdot \int f_\ell \, d\mu \tag{13}$$

where the convergence takes place in $L^2(\mu)$.

[5] A comment about notation. In [7] the class $\mathcal{LE}$ is denoted by $\mathcal{H}$. Also, for functions in $\mathcal{LE}$ our class $\mathcal{G}$ coincides with the class $\mathcal{T}$ defined in [7].

Remarks. • Some examples for which our result applies are given by the collections of sequences $\{[n^{1/2}],[n^{3/2}],[n^{5/2}]\}$, and $\{[n^{\sqrt{2}}],[n^{\sqrt{2}}\log\log n],[n^{\sqrt{2}}\log n]\}$.

• Equation (13) fails for some ergodic systems if a non-trivial linear combination of the functions $a_1(t),\ldots,a_\ell(t)$ is an integer polynomial other than $\pm t+k$.

• A substantial part of the proof (carried out in the companion paper [14]) is consumed in working on a potentially non-trivial (characteristic) factor of our system. Initially we show that this factor has (roughly speaking) the structure of a nilsystem, only to realize later (using some non-trivial equidistribution results on nilmanifolds) that this factor is trivial. It would be nice to have a proof that avoids such diversions to non-Abelian analysis.

If all the functions $a_1(t),\ldots,a_\ell(t)$ have sub-linear growth then Theorem 2.6 can be (rather easily) proved in a more general setup, where one uses iterates of ℓ not necessarily commuting ergodic transformations in place of a single ergodic transformation.

Theorem 2.7. *Let $a_1,\ldots,a_\ell\in\mathcal{LE}\cap\mathcal{G}$ have different growth rates and satisfy $a_i(t)\prec t$ for $i=1,\ldots,\ell$. Let $T_1,\ldots,T_\ell$ be invertible measure preserving transformations acting on a probability space $(X,\mathcal{X},\mu)$.*

Then for every $f_1,\ldots,f_\ell\in L^\infty(\mu)$ we have

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N} f_1(T_1^{[a_1(n)]}x)\cdot\ldots\cdot f_\ell(T_\ell^{[a_\ell(n)]}x)=\tilde{f}_1\cdot\ldots\cdot\tilde{f}_\ell,$$

where $\tilde{f}_i=\mathbb{E}(f_i|\mathcal{I}(T_i))$, and the convergence takes place in $L^2(\mu)$.

Remarks. • In [7] a similar result was proved for iterates of a single transformation.

• It is not known whether similar convergence results hold without any commutativity assumption on the transformations T_i for some choice of functions $a_i(t)$ with different, at least linear growth rates. On the other hand, it is known ([2]) that for some choice of non-commuting transformations T_1,T_2 and functions f_1,f_2, the averages $\frac{1}{N}\sum_{n=1}^{N}f_1(T_1^n x)\cdot f_2(T_2^n x)$ diverge in $L^2(\mu)$.

2.2.2. *Recurrence.* The next multiple recurrence result is a consequence of Theorem 2.6:

Theorem 2.8. *Suppose that the functions $a_1,\ldots,a_\ell\in\mathcal{LE}\cap\mathcal{G}$ have different growth rates.*

Then for every system $(X,\mathcal{X},\mu,T)$ and set $A\in\mathcal{X}$ we have

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}\mu(A\cap T^{-[a_1(n)]}A\cap T^{-[a_2(n)]}A\cap\cdots\cap T^{-[a_\ell(n)]}A)\geq(\mu(A))^{\ell+1}. \tag{14}$$

Remarks. • The estimate (14) becomes an equality when the system is ergodic.

• The lower bounds (14) contrast the corresponding lower bounds when the functions $a_1(t),...,a_\ell(t)$ are non-constant integer polynomials. In this case, (14) fails even when $\ell=1$ and $a_1(t)=t^2$. In fact no power type lower bound is known for any collection of polynomials (except of course when all the functions are equal and linear).

2.2.3. *Characteristic factors.* The next result gives convenient characteristic factors for a family of "typical" logarithmico-exponential sequences of polynomial growth.

Theorem 2.9. *Let $a_1,\ldots,a_\ell\in\mathcal{LE}$, and suppose that all the functions $a_i(t)$ and their pairwise differences $a_i(t)-a_j(t)$ belong in $\mathcal{G}$ (defined (12)).*

Then for every system its nilfactor $\mathcal{Z}$ is characteristic for the family $\{[a_1(n)],\ldots,[a_\ell(n)]\}$.

Remark. A related result was proved in [7] for weakly mixing systems. In fact we are going to adapt the argument used in [7] to establish our result.

2.2.4. *Combinatorics.* Using Furstenberg's Correspondence Principle and Theorem 2.8 we immediately deduce the following:

Theorem 2.10. *Suppose that the functions $a_1,\ldots,a_\ell \in \mathcal{LE}\cap\mathcal{G}$ have different growth rates. Then for every set of integers Λ we have*

$$\liminf_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}\overline{d}(\Lambda\cap(\Lambda-[a_1(n)])\cap\cdots\cap(\Lambda-[a_\ell(n)]))\geq\left(\overline{d}(\Lambda)\right)^{\ell+1}.$$

2.3. **Further directions.** We state some open problems that are closely related to the results stated before. To avoid repetition we remark that *in Problems 1-4 we always work with a family $\mathcal{F}=\{a_1(t),\ldots,a_\ell(t)\}$ of functions of polynomial growth that belong to the same Hardy field*. With span($\mathcal{F}$) we denote the set of all *non-trivial* linear combinations of elements of $\mathcal{F}$.

2.3.1. *Convergence.* The family of functions $\mathcal{F}=\{a_1(t),\ldots,a_\ell(t)\}$ is *good for multiple convergence* if for every system $(X,\mathcal{X},\mu,T)$ and functions $f_1,\ldots,f_\ell\in L^\infty(\mu)$ the limit

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}f_1(T^{[a_1(n)]}x)\cdot\ldots\cdot f_\ell(T^{[a_\ell(n)]}x) \tag{15}$$

exists in $L^2(\mu)$.

The next problem is much in the spirit of Theorem 2.1:

Problem 1. *The family $\mathcal{F}$ is good for multiple convergence if and only if every function $a\in$ span($\mathcal{F}$) satisfies one of the following conditions:*

- $|a(t)-cp(t)|\succ\log t$ *for every $c\in\mathbb{R}$ and every $p\in\mathbb{Z}[t]$; or*
- $a(t)-cp(t)\to d$ *for some $c,d\in\mathbb{R}$; or*
- $|a(t)-t/m|\ll\log t$ *for some $m\in\mathbb{Z}$.*

The next problem provides a possible generalization of Theorem 2.6:

Problem 2. *Suppose that every function $a\in$ span($\mathcal{F}$) satisfies $|a(t)-cp(t)|\succ\log t$ for every $c\in\mathbb{R}$ and every $p\in\mathbb{Z}[t]$.*

Then for every ergodic system $(X,\mathcal{B},\mu,T)$ and $f_1,\ldots,f_\ell\in L^\infty(\mu)$ we have

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}f_1(T^{[a_1(n)]}x)\cdot\ldots\cdot f_\ell(T^{[a_\ell(n)]}x)=\int f_1\,d\mu\cdot\ldots\cdot\int f_\ell\,d\mu \tag{16}$$

where the convergence takes place in $L^2(\mu)$.

We remark that if some function $a\in$ span($\mathcal{F}$) satisfies $|a(t)-cp(t)|\ll\log t$ for some $c\in\mathbb{R}$ and $p\in\mathbb{Z}[t]$ with $\deg(p)\geq 2$, then (16) fails for some system.

2.3.2. *Characteristic factors.* We state a possible generalization of Theorem 2.9:

Problem 3. *Suppose that $a_i(t)\succ\log t$ and $a_i(t)-a_j(t)\succ\log t$ whenever $i\neq j$.*

Then for every system its nilfactor $\mathcal{Z}$ is characteristic for the family $\{[a_1(n)],\ldots,[a_\ell(n)]\}$.

One can easily see that the stated assumptions are also necessary.

2.3.3. *Recurrence.* The next problem provides a possible extension of Theorem 2.3:

Problem 4. *Suppose that every function $a \in span(\mathcal{F})$ satisfies $|a(t) - cp(t)| \to \infty$ for every $c \in \mathbb{R}$ and every $p \in \mathbb{Z}[t]$.*

Then for every system $(X, \mathcal{X}, \mu, T)$ and $A \in \mathcal{X}$ with $\mu(A) > 0$ we have

$$\mu(A \cap T^{-[a_1(n)]}A \cap \cdots \cap T^{-[a_\ell(n)]}A) > 0$$

for some $n \in \mathbb{N}$ such that $[a_i(n)] \neq 0$.

An interesting special case of this result is when the functions $a_1(t), \ldots, a_\ell(t)$ have different growth and none of them is equal to a polynomial (modulo a function that vanishes at infinity).

If all the functions $a_1(t), \ldots, a_\ell(t)$ are integer polynomials, then necessary and sufficient conditions for multiple recurrence where given in [10].

2.3.4. *Combinatorics.* We rephrase Problem 4 in combinatorial terminology:

Problem 4′. *Suppose that every function $a \in span(\mathcal{F})$ satisfies $|a(t) - cp(t)| \to \infty$ for every $c \in \mathbb{R}$ and every $p \in \mathbb{Z}[t]$.*

Then every $\Lambda \subset \mathbb{Z}$ with $\bar{d}(\Lambda) > 0$ contains patterns of the form

$$\{m, m + [a_1(n)], \ldots, m + [a_\ell(n)]\} \tag{17}$$

for some $m \in \mathbb{Z}$ and $n \in \mathbb{N}$ with $[a_i(n)] \neq 0$.

2.3.5. *Commuting transformations.* It seems likely that our main results remain true when one works with iterates of ℓ commuting measure preserving transformations instead of iterates a single transformation. We state two related problems here:

Problem 5. *Let $T_1, \ldots, T_\ell$ be commuting invertible measure preserving transformations acting on a probability space $(X, \mathcal{X}, \mu)$ and $f_1, \ldots, f_\ell \in L^\infty(\mu)$.*

Then for every positive real number c the following limit exists in $L^2(\mu)$

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} f_1(T_1^{[n^c]}x) \cdot \ldots \cdot f_\ell(T_\ell^{[n^c]}x). \tag{18}$$

Furthermore, if c is not an integer, then (18) *is equal to* $\lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} T_1^n f_1 \cdot \ldots \cdot T_\ell^n f_\ell$.

For $c = 1$ the existence of the limit (18) was established by Tao in [33] (see also [35],[1], [25] for other subsequent proofs). The case $0 < c < 1$ can be easily reduced to the case $c = 1$ (see Lemma 5.1 below).

Problem 6. *Let $T_1, \ldots, T_\ell$ be commuting measure preserving transformations acting on a probability space $(X, \mathcal{X}, \mu)$. Let $c_1, \ldots, c_\ell \in \mathbb{R} \setminus \mathbb{Z}$ be positive and distinct.*

Then for every $f_1, \ldots, f_\ell \in L^\infty(\mu)$ we have

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} f_1(T_1^{[n^{c_1}]}x) \cdot \ldots \cdot f_\ell(T_\ell^{[n^{c_\ell}]}x) = \tilde{f}_1 \cdot \ldots \cdot \tilde{f}_\ell,$$

where $\tilde{f}_i = \mathbb{E}(f_i|\mathcal{I}(T_i))$, and the convergence takes place in $L^2(\mu)$.

This is an immediate consequence of Theorem 2.7 when all the exponents c_i are smaller than 1 and the commutativity of the transformations T_i is not needed in this case.

2.3.6. *Prime numbers.* The results of this article should remain true when one makes the substitution $n \to p_n$ where p_n denotes the n-th prime number. For instance:

Problem 7. *If c is a positive non-integral real number, then the sequence $[p_n^c]$ is good for multiple recurrence and convergence. Furthermore, the limit of the corresponding multiple ergodic averages is equal to the limit of the "Furstenberg averages" (defined by* (3)*).*

One could try to verify such a statement by comparing the averages along $[p_n^c]$ to the averages along the sequence $[n^c]$ for which all required properties are known. A similar strategy was used in [15] to deal with double recurrence (and convergence) problems of the shifted primes.

Another challenge is to use the Szemerédi type results of Sections 2.1 and 2.2 and prove that the primes contain the corresponding Hardy-field patterns. For instance:

Problem 8. *If $c, c_1, c_2 \in \mathbb{R}$ are positive, then the prime numbers contain patterns of the form*

$$\{m, m + [n^c], m + 2[n^c]\} \quad \textit{and} \quad \{m, m + [n^{c_1}], m + [n^{c_2}]\}.$$

We remark that using the corresponding density results (in addition to many other things), Green and Tao ([21]) proved the existence of arbitrarily long arithmetic progressions in the primes, and Tao and Ziegler ([34]) the existence of arbitrarily long polynomial progressions in the primes (this last result allows one to handle Problem 8 when c, c_1, c_2 are positive rational numbers).

2.4. **Structure of the article and main ideas.** In Section 3 we gather some essential background material from ergodic theory, equidistribution results on nilmanifolds, and basic facts about Hardy fields. Key for our study is the structure theorem of Host and Kra (Theorem 3.3) and the quantitative equidistribution result of Green and Tao (Theorem 3.2). The use of the latter result is rather implicit in this article since we frequently use results from the companion article [14] that were proved using quantitative equidistribution.

In Section 4 we prove Theorem 2.4 which shows that under appropriate conditions the nilfactor is characteristic for families of the form $\{[a(n)], \ldots, \ell[a(n)]\}$. We remark that for functions of "fractional-power" growth (like $a(t) = t^{3/2}$), this problem can be handled using more or less conventional techniques. But for functions that have slowly growing derivatives (like $a(t) = t \log t$) the "standard" techniques become problematic. To overcome this problem, we partition the positive integers into intervals of appropriate size, and in each such interval we use the Taylor expansion of the function to get an approximation by real valued polynomials of fixed degree. This approximation works well when the function stays logarithmically away from polynomials, and as a result functions like $t^{3/2}$, $t \log t$, and $t + (\log t)^2$ become practically indistinguishable for our purposes. After performing these initial maneuvers we are led to estimating some multiple ergodic averages involving polynomial iterates (Proposition 4.1), a problem that can be handled using more or less standard techniques.

In Section 5 we prove Theorem 2.9 which shows that under appropriate conditions the nilfactor is characteristic for families of the form $\{[a_1(n)], \ldots, [a_\ell(n)]\}$. Since we only work with functions of "fractional-power" growth, we are able to adapt an argument of Bergelson and Håland ([7]) that was used to establish a convergence result for weakly mixing systems. The proof consists of two steps. One first deals with the case where all the functions have at most linear growth (Proposition 5.3). This is done by successively applying Van der Corput's lemma and a change of variable trick. Then one uses a modification of the polynomial exhaustion technique of Bergelson to reduce the general case to the case of at most linear growth.

In the last section we complete the proof of the convergence and recurrence results of Sections 2.1 and 2.2. With the exception of Theorem 2.7 that can be handled directly, to prove the other convergence results we first make use of the results from Sections 4 and 5 to show that the nilfactor of the system is characteristic for the appropriate multiple ergodic averages. Then Theorem 3.3 enables us to reduce matters to nilsystems. Finally, we use equidistribution results from the companion paper [14] to verify the appropriate convergence property for nilsystems. The recurrence results are direct consequences of the corresponding convergence results, with the exception of a special case of Theorem 2.3 where the function is logarithmically close to a constant multiple of an integer polynomial. In this case, a somewhat complicated analysis is used to prove the corresponding recurrence property for nilsystems.

2.5. **Notational conventions.** The following notation will be used throughout the article: $\mathbb{N} = \{1, 2, \ldots\}$, $\mathbb{T}^k = \mathbb{R}^k/\mathbb{Z}^k$, $Tf = f \circ T$, $e(t) = e^{2\pi i t}$, $[t]$ denotes the integer part of t, $\{t\} = t - [t]$, $\|x\| = d(t, \mathbb{Z})$, $\mathbb{E}_{n\in A} a(n) = \frac{1}{|A|}\sum_{n\in A} a(n)$. We sometimes write t to represent an element $t\mathbb{Z}$ of $\mathbb{T} = \mathbb{R}/\mathbb{Z}$. By $a(t) \prec b(t)$ we mean $\lim_{t\to\infty} a(t)/b(t) = 0$, and by $a(t) \sim b(t)$ we mean $\lim_{t\to\infty} a(t)/b(t)$ is a non-zero real number, and by $a(t) \ll b(t)$ we mean $|a(t)| \leq C|b(t)|$ for some absolute constant C. By $\mathbb{R}_+$ we denote some half-line $(c, +\infty)$. When often write ∞ instead of $+\infty$. By $\mathcal{H}$ we denote the class of all functions that belong to some Hardy field, by $\mathcal{LE}$ the class of all logarithmico-exponential functions, and by $\mathcal{G}$ the class of functions in $C(\mathbb{R}_+)$ that satisfy $t^{k+\varepsilon} \prec a(t) \prec t^{k+1}$ for some non-negative integer k and $\varepsilon > 0$.

3. Background material

3.1. **Hardy fields.** We collect here some basic properties of elements of $\mathcal{H}$ relevant to our study. The reader can find more background material in [11] and the references therein.

Every element of $\mathcal{H}$ has eventually constant sign. Therefore, if $a \in \mathcal{H}$, then $a(t)$ is eventually monotonic (since $a'(t)$ has eventually constant sign), and the limit $\lim_{t\to\infty} a(t)$ exists (possibly infinite). For every two functions $a \in \mathcal{H}, b \in \mathcal{LE}$ ($b \neq 0$), we have $a/b \in \mathcal{H}$. It follows that the asymptotic growth ratio $\lim_{t\to\infty} a(t)/b(t)$ exists (possibly infinite).

We caution the reader that $\mathcal{H}$ is not a field, and some pairs of functions in $\mathcal{H}$ are not asymptotically comparable. This defect of $\mathcal{H}$ plays a role in some of our results, and can be sidestepped by restricting our attention to the Hardy field of logarithmico-exponential functions $\mathcal{LE}$.

A key property of elements of $\mathcal{H}$ with polynomial growth is that one can relate their growth rates with the growth rates of their derivatives:

Lemma 3.1 (**Lemma 2.1 in** [14]). *Suppose that $a \in \mathcal{H}$ has polynomial growth. We have the following*

(i) If $t^\varepsilon \prec a(t)$ for some $\varepsilon > 0$, then $a'(t) \sim a(t)/t$.

(ii) If $t^{-k} \prec a(t)$ for some $k \in \mathbb{N}$, and $a(t)$ does not converge to a non-zero constant, then $a(t)/(t(\log t)^2) \prec a'(t) \ll a(t)/t$.

We are going to freely use all these properties in the sequel.

3.2. **Nilmanifolds.** All the equidistribution results needed in this article were established in [14] with the exception of one result that will be needed to cover a special case of Theorem 2.3. Below we gather some basic facts and a quantitative equidistribution result that will be used in its proof.

The proofs of all the results mentioned below can be found or deduced from [28] and [22].

3.2.1. *Definitions and basic properties.* A *nilmanifold* is a homogeneous space $X = G/\Gamma$ where G is a nilpotent Lie group, and Γ is a discrete cocompact subgroup of G. If $G_{k+1} = \{e\}$, where G_k denotes the k-th commutator subgroup of G, we say that X is a *k-step* nilmanifold. With G_0 we denote the connected component of the identity element in G. The representation of a nilmanifold X as a homogeneous space of a nilpotent Lie group G is not unique. It can be shown ([28]) that if X is connected, then it admits a representation of the form $X = G/\Gamma$ such that G_0 is simply connected and $G = G_0\Gamma$. *For connected nilmanifolds $X = G/\Gamma$, we will always assume that G satisfies these two extra assumptions.*

The group G acts on G/Γ by left translation where the translation by a fixed element $b \in G$ is given by $T_b(g\Gamma) = (bg)\Gamma$. By m_X we denote the unique probability measure on X that is invariant under the action of G by left translations (called the *normalized Haar measure*) and $\mathcal{G}/\Gamma$ denote the Borel σ-algebra of G/Γ. Fixing an element $b \in G$, we call the system $(G/\Gamma, \mathcal{G}/\Gamma, m, T_b)$ a *nilsystem*. We call the elements of G *nilrotations*.

For every $b \in G$ the set $X_b = \overline{\{b^n\Gamma, n \in \mathbb{N}\}}$ is a nilmanifold H/Δ, where H is a closed subgroup of G that contains b, and $\Delta = H \cap \Gamma$ is a discrete cocompact subgroup of H. Furthermore, for every $b \in G$ there exists an $r \in \mathbb{N}$ such that the nilmanifold X_{b^r} is connected.

A nilrotation $b \in G$ is *ergodic*, or *acts ergodically on X*, if the sequence $(b^n\Gamma)_{n\in\mathbb{N}}$ is dense in X. If $b \in G$ is ergodic, then for every $x \in X$ the sequence $(b^n x)_{n\in\mathbb{N}}$ is equidistributed in X. If the nilmanifold X is connected and b acts ergodically on X, then for every $r \in \mathbb{N}$ the element b^r also acts ergodically on X.

3.2.2. *A quantitative equidistribution result.* If G is a nilpotent group, then a sequence $g\colon \mathbb{Z} \to G$ of the form $g(n) = b_1^{p_1(n)} \cdot \ldots \cdot b_k^{p_k(n)}$, where $b_i \in G$, and p_i are polynomials taking integer values at the integers, is called a *polynomial sequence in G*. If the maximum degree of the polynomials p_i is at most d we say that the *degree* of $g(n)$ is at most d. A *polynomial sequence on the nilmanifold $X = G/\Gamma$* is a sequence of the form $(g(n)\Gamma)_{n\in\mathbb{Z}}$ where $g\colon \mathbb{Z} \to G$ is a polynomial sequence in G.

In [22], Green and Tao proved a quantitative equidistribution result for polynomial sequences on nilmanifolds $X = G/\Gamma$ when the group G is connected and simply connected. We will need an extension of this result to the non-connected case. In order to state it we first introduce some notions from [22] and [17].

If $X = G/\Gamma$ is a connected nilmanifold, the *affine torus* of X is defined to be the homogeneous space $A = G/([G_0, G_0]\Gamma)$. It is known ([16]) that every nilrotation acting on the affine torus is isomorphic to a unipotent affine transformation on some finite dimensional torus[6] with the normalized Haar measure, and furthermore the conjugation map can be taken to be continuous. We can therefore identify the affine torus A of a nilmanifold X with a finite dimensional torus $\mathbb{T}^l$ and think of a nilrotation acting on A as a unipotent affine transformation on $\mathbb{T}^l$.

A *quasi-character* of a nilmanifold $X = G/\Gamma$ is a function $\psi\colon G \to \mathbb{C}$ that is a continuous homomorphism of G_0 (to the multiplicative group $\{z \in \mathbb{C}\colon |z| = 1\}$) and satisfies $\psi(g\gamma) = \psi(g)$ for every $\gamma \in \Gamma$. Every quasi-character annihilates $[G_0, G_0]$, and as a result factors through the affine torus A of X. Under an appropriate isomorphism we have that $A \simeq \mathbb{T}^l$ and every quasi-character of X is mapped to a character of $\mathbb{T}^l$. Therefore, thinking of ψ as a character of $\mathbb{T}^l$ we have $\psi(t) = e(\kappa \cdot t)$ for some $\kappa \in \mathbb{Z}^l$, where $\cdot$ denotes the inner product operation. We refer to κ as the *frequency* of ψ and $\|\psi\| = |\kappa|$ as the *frequency magnitude* of ψ.

[6]This means $T\colon \mathbb{T}^l \to \mathbb{T}^l$ has the form $T(t) = b \cdot S(t)$ for some unipotent homomorphism S of $\mathbb{T}^l$ and $b \in \mathbb{T}^l$.

If $p\colon \mathbb{Z} \to \mathbb{R}$ is a polynomial sequence of degree k, then p can be uniquely expressed in the form $p(n) = \sum_{i=0}^{k} \binom{n}{i} \alpha_i$ where $\alpha_i \in \mathbb{R}$. For $N \in \mathbb{N}$ we define

$$\|e(p(n))\|_{C^\infty[N]} = \max_{1\le i\le k} (N^i \|\alpha_i\|) \tag{19}$$

where $\|t\| = d(t, \mathbb{Z})$.

Given $N \in \mathbb{N}$, a finite sequence $(g(n)\Gamma)_{1\le n\le N}$ is said to be δ-*equidistributed* if

$$\Big|\frac{1}{N}\sum_{n=1}^{N} F(g(n)\Gamma) - \int_X F\ dm_X\Big| \le \delta \|F\|_{\mathrm{Lip}(X)}$$

for every Lipschitz function $F\colon X \to \mathbb{C}$ where $\|F\|_{\mathrm{Lip}(X)} = \|F\|_\infty + \sup_{x,y\in X, x\ne y} \frac{|F(x)-F(y)|}{d_X(x,y)}$ for some appropriate metric d_X on X.

The next result can be found in [17] (see Theorem 3.4).

Theorem 3.2 (**Corollary of Green & Tao** [22])**.** *Let $X = G/\Gamma$ be a connected nilmanifold (with G_0 simply connected), and $d \in \mathbb{N}$.*

Then for every small enough $\delta > 0$ there exists $M = M_{X,d,\delta} \in \mathbb{R}$ with the following property: For every $N \in \mathbb{N}$, if $g\colon \mathbb{Z} \to G$ is a polynomial sequence of degree at most d such that the finite sequence $(g(n)\Gamma)_{1\le n\le N}$ is not δ-equidistributed, then for some non-trivial quasi-character ψ with $\|\psi\| \le M$ we have

$$\|\psi(g(n))\|_{C^\infty[N]} \le M$$

where we think of ψ as a character of some finite dimensional torus $\mathbb{T}^l$ (the affine torus) and $g(n)$ as a polynomial sequence of unipotent affine transformations on $\mathbb{T}^l$.

Remark. We have $\psi(g(n)) = e(p(n))$ for some $p \in \mathbb{R}[x]$ and therefore $\|\psi(g(n))\|_{C^\infty[N]}$ is well defined.

3.3. **Ergodic theory.** Below we gather some basic notions and facts from ergodic theory that we use throughout the paper. The reader can find further background material in ergodic theory in [19], [31], [36].

3.3.1. *Factors.* A *homomorphism* from a system $(X, \mathcal{X}, \mu, T)$ onto a system $(Y, \mathcal{Y}, \nu, S)$ is a measurable map $\pi\colon X' \to Y'$, where X' is a T-invariant subset of X and Y' is an S-invariant subset of Y, both of full measure, such that $\mu \circ \pi^{-1} = \nu$ and $S \circ \pi(x) = \pi \circ T(x)$ for $x \in X'$. When we have such a homomorphism we say that the system $(Y, \mathcal{Y}, \nu, S)$ is a *factor* of the system $(X, \mathcal{X}, \mu, T)$. If the factor map $\pi\colon X' \to Y'$ can be chosen to be injective, then we say that the systems $(X, \mathcal{X}, \mu, T)$ and $(Y, \mathcal{Y}, \nu, S)$ are *isomorphic* (bijective maps on Lebesgue spaces have measurable inverses).

A factor can be characterized (modulo isomorphism) by $\pi^{-1}(\mathcal{Y})$ which is a T-invariant sub-σ-algebra of $\mathcal{B}$, and conversely any T-invariant sub-σ-algebra of $\mathcal{B}$ defines a factor. By a classical abuse of terminology we denote by the same letter the σ-algebra $\mathcal{Y}$ and its inverse image by π. In other words, if $(Y, \mathcal{Y}, \nu, S)$ is a factor of $(X, \mathcal{X}, \mu, T)$, we think of $\mathcal{Y}$ as a sub-σ-algebra of $\mathcal{X}$. A factor can also be characterized (modulo isomorphism) by a T-invariant subalgebra $\mathcal{F}$ of $L^\infty(X, \mathcal{X}, \mu)$, in which case $\mathcal{Y}$ is the sub-σ-algebra generated by $\mathcal{F}$, or equivalently, $L^2(X, \mathcal{Y}, \mu)$ is the closure of $\mathcal{F}$ in $L^2(X, \mathcal{X}, \mu)$. We will sometimes abuse notation and use the sub-σ-algebra $\mathcal{Y}$ in place of the subspace $L^2(X, \mathcal{Y}, \mu)$. For example, if we write that a function is orthogonal to the factor $\mathcal{Y}$, we mean it is orthogonal to the subspace $L^2(X, \mathcal{Y}, \mu)$.

If $\mathcal{Y}$ is a T-invariant sub-σ-algebra of $\mathcal{X}$ and $f \in L^2(\mu)$, we define the *conditional expectation* $\mathbb{E}(f|\mathcal{Y})$ *of* f *with respect to* $\mathcal{Y}$ to be the orthogonal projection of f onto $L^2(\mathcal{Y})$. We will frequently make use of the identities

$$\int \mathbb{E}(f|\mathcal{Y})\, d\mu = \int f\, d\mu, \quad T\,\mathbb{E}(f|\mathcal{Y}) = \mathbb{E}(Tf|\mathcal{Y}).$$

If we want to indicate the dependence on the reference measure, we write $\mathbb{E} = \mathbb{E}_\mu$.

The transformation T is *ergodic* if $Tf = f$ implies that $f = c$ (a.e.) for some $c \in \mathbb{C}$. Every system $(X, \mathcal{X}, \mu, T)$ has an *ergodic decomposition*, meaning that we can write $\mu = \int \mu_t\, d\lambda(t)$, where λ is a probability measure on $[0, 1]$ and μ_t are T-invariant probability measures on $(X, \mathcal{X})$ such that the systems $(X, \mathcal{X}, \mu_t, T)$ are ergodic for $t \in [0, 1]$.

We say that $(X, \mathcal{X}, \mu, T)$ is an *inverse limit of a sequence of factors* $(X, \mathcal{X}_j, \mu, T)$ if $(\mathcal{X}_j)_{j\in\mathbb{N}}$ is an increasing sequence of T-invariant sub-σ-algebras such that $\bigvee_{j\in\mathbb{N}} \mathcal{X}_j = \mathcal{X}$ up to sets of measure zero.

3.3.2. *Seminorms and nilfactors.* Following [26],[7] for every system $(X, \mathcal{X}, \mu, T)$ and function $f \in L^\infty(\mu)$, we define inductively the seminorms $\|\!| f |\!\|_\ell$ as follows: For $\ell = 1$ we set

$$\|\!| f |\!\|_1 = \|\mathbb{E}(f|\mathcal{I})\|_{L^2(\mu)}$$

where $\mathcal{I}$ is the σ-algebra of T-invariant sets. For $\ell \geq 1$ we set

$$\|\!| f |\!\|_{\ell+1}^{2^{\ell+1}} = \lim_{N\to\infty} \mathbb{E}_{1\leq n\leq N} \|\!| \bar{f} \cdot T^n f |\!\|_\ell^{2^\ell}. \tag{20}$$

[8]

It was shown in [26] that for every integer $\ell \geq 1$, $\|\!| \cdot |\!\|_\ell$ is a seminorm on $L^\infty(\mu)$ and it defines factors $\mathcal{Z}_{\ell-1} = \mathcal{Z}_{\ell-1}(T)$ in the following manner: the T-invariant sub-σ-algebra $\mathcal{Z}_{\ell-1}$ is characterized by

$$\text{for } f \in L^\infty(\mu),\ \mathbb{E}(f|\mathcal{Z}_{\ell-1}) = 0 \text{ if and only if } \|\!| f |\!\|_\ell = 0.$$

It is shown in [26] that for every $\ell \in \mathbb{N}$ the factor $\mathcal{Z}_\ell$ has a purely algebraic structure, in fact for all practical purposes we can assume that it is an ℓ-step nilsystem.

Theorem 3.3 (Host & Kra [26]**).** *Let $(X, \mathcal{X}, \mu, T)$ be a system and $\ell \in \mathbb{N}$.*
Then a.e. ergodic component of the factor $\mathcal{Z}_\ell(T)$ is an inverse limit of ℓ-step nilsystems.

Because of this result we call $\mathcal{Z}_\ell$ the *ℓ-step nilfactor* of the system. The smallest factor that is an extension of all finite step nilfactors is denoted by $\mathcal{Z}$ and is called the *nilfactor* of the system (in other words $\mathcal{Z} = \bigvee_{j\in\mathbb{N}} \mathcal{Z}_j$.) The nilfactor $\mathcal{Z}$ is of particular interest because, as it turns out, it controls the limiting behavior in $L^2(\mu)$ of the multiple ergodic averages that are studied in Theorems 2.1 and 2.6.

We also record two useful identities that can be easily established using the induction definition of the seminorms

$$\|\!| f |\!\|_\ell^{2^\ell} = \int \|\!| f |\!\|_{\mu_s,\ell}^{2^\ell}\, d\lambda(s), \qquad \|\!| f |\!\|_{\ell+1}^2 = \|\!| f \otimes \overline{f} |\!\|_{\mu\times\mu,\ell}. \tag{21}$$

where $\mu = \int \mu_s\, d\lambda(s)$ is the ergodic decomposition associated to the system $(X, \mathcal{X}, \mu, T)$. Hence, if T_t where $t \in [0, 1]$ are the ergodic components of the transformation T, then $\mathbb{E}(f|\mathcal{Z}_\ell(T)) = 0$

[7] In [26] the authors work with ergodic systems and real valued functions, but the whole discussion can be carried out for non-ergodic systems as well and complex valued functions without extra difficulties.

[8] We remark that the limit is the same if the average $\mathbb{E}_{1\leq n\leq N}$ is replaced with the average $\mathbb{E}_{n\in\Phi_N}$ where $(\Phi_N)_{N\in\mathbb{N}}$ is any Følner sequence in $\mathbb{Z}$.

if and only if $\mathbb{E}(f|\mathcal{Z}_\ell(T_t)) = 0$ for a.e. $t \in [0,1]$. Also if f satisfies $\mathbb{E}_\mu(f|\mathcal{Z}_\ell(T)) = 0$, then $\mathbb{E}_{\mu\times\mu}(f \otimes \overline{f}|\mathcal{Z}_{\ell-1}(T\times T)) = 0$.

4. Characteristic factors for multiples of a single sequence

A crucial step in the proof of Theorem 2.1 is to show that for every $a \in \mathcal{H}$, not growing very fast or very slow, for every $\ell \in \mathbb{N}$, the nilfactor $\mathcal{Z}$ of a system is characteristic for the family $\{[a(n)], 2[a(n)], \dots, \ell[a(n)]\}$. This is the context of Theorem 2.4 which we are going to prove in this section.

As is typically the case with such results, one assigns a notion of "complexity" to the relevant multiple ergodic averages, and then uses induction on the "complexity" to prove the result. This plan can be carried out without serious difficulties when the function $a \in \mathcal{H}$ satisfies $t^{k-1}\log t \prec a(t) \prec t^k$ for some $k \in \mathbb{N}$. But when $a(t) = t\log t$, for example, there are serious difficulties caused by the fact that the factor $\mathcal{Z}$ is not characteristic for the Hardy sequence $[a'(n)]$, the reason being that the sequence $[\log n]$ grows too slowly. To deal with such functions we perform some initial maneuvers that enable us to transform the problem to one where induction on the "complexity" is applicable. Before giving the formal argument we informally explain how the initial step of the proof works in a model case.

4.1. **A model problem.** Suppose we want to show that the nilfactor $\mathcal{Z}$ is characteristic for the family of sequences $\{[n\log n], 2[n\log n], \dots, \ell[n\log n]\}$. Let $(X, \mathcal{X}, \mu, T)$ be a system, and suppose that one of the functions $f_1, \dots, f_\ell \in L^\infty(\mu)$ is orthogonal to $\mathcal{Z}$. We have to show that

$$\lim_{N\to\infty} \mathbb{E}_{1\leq n\leq N} V([n\log n]) = 0, \tag{22}$$

where

$$V(n) = T^n f_1 \cdot T^{2n} f_2 \cdot \ldots \cdot T^{\ell n} f_\ell,$$

and the convergence takes place in $L^2(\mu)$. Our goal here is to show how to transform (22) into a more manageable identity.

It will be more convenient for us to show that

$$\lim_{N\to\infty} \mathbb{E}_{N<n\leq N+l(N)} V([n\log n]) = 0 \tag{23}$$

for some function $l(t)$ that satisfies $l(t) \prec t$ (Lemma 4.3 shows that (23) implies (22)). Using the Taylor expansion of $a(t) = t\log t$ around the point $t = N$ we get for every $n \in \mathbb{N}$ that

$$(N+n)\log(N+n) = N\log N + n(1+\log N) + \frac{n^2}{2N} - \frac{n^3}{6\xi_n^2}$$

for some $\xi_n \in [N, N+n]$. Hence, if $c < 2/3$, then for every large N and $n = 1, \dots, [N^c]$, we have

$$[(N+n)\log(N+n)] = \Big[\alpha_N + n\beta_N + \frac{n^2}{2N}\Big] + e(n)$$

for some $\alpha_N, \beta_N \in \mathbb{R}$ and error terms $e(n) \in \{0, -1\}$. Ignoring the error terms, and writing $[\sqrt{2N}]^2$ in place of $2N$ (all these technical issues can be justified), we get that in order to establish (23) it suffices to show that

$$\lim_{N\to\infty} \mathbb{E}_{1\leq n\leq N^c} V\Big(\Big[\alpha_N + n\beta_N + \frac{n^2}{[\sqrt{2N}]^2}\Big]\Big) = 0. \tag{24}$$

Since every integer between 1 and N^c can be represented as $n[\sqrt{2N}]+r$ with $0\leq n\leq \tilde{l}(N)=N^c/[\sqrt{2N}]$ and $1\leq r\leq r(N)=[\sqrt{2N}]$, and since

$$\frac{(n[\sqrt{2N}]+r)^2}{[\sqrt{2N}]^2}+(n[\sqrt{2N}]+r)\alpha+\beta=n^2+n\alpha_{r,N}+\beta_{r,N}$$

for some $\alpha_{r,N},\beta_{r,N}\in\mathbb{R}$, we get that (24) follows if we show that

$$\lim_{N\to\infty}\mathbb{E}_{1\leq r\leq r(N)}\Big(\mathbb{E}_{1\leq n\leq \tilde{l}(N)}V([\alpha_{r,N}+n\beta_{r,N}+n^2])\Big)=0.$$

If we choose $c>1/2$, then $\tilde{l}(N)\to\infty$, and as a result the last identity follows if we show that

$$\lim_{N\to\infty}\sup_{\alpha,\beta\in\mathbb{R}}\big\|\mathbb{E}_{1\leq n\leq N}V([\alpha+n\beta+n^2])\big\|_{L^2(\mu)}=0. \tag{25}$$

We have therefore reduced matters to establishing uniform estimates for some polynomial multiple ergodic averages, and this turns out to be a more manageable problem.

We also remark that the argument used in the previous model example turns out to work for every function $a\in\mathcal{H}$ of polynomial growth that satisfies $|a(t)-p(t)|\succ\log t$ for every $p\in\mathbb{R}[t]$. We give the details in the next subsection.

4.2. **Proof of Theorem 2.4 modulo a polynomial ergodic theorem.** We are going to prove Theorem 2.4 modulo the following polynomial ergodic theorem that we shall prove in the next subsection:

Proposition 4.1. *Let $(X,\mathcal{X},\mu,T)$ be a system, and suppose that at least one of the functions $f_1,f_2,\ldots,f_\ell\in L^\infty(\mu)$ is orthogonal to the nilfactor $\mathcal{Z}$.*

Then for every $k\in\mathbb{N}$, nonzero $\alpha\in\mathbb{R}$, bounded two parameter sequence $(c_{N,n})_{N,n\in\mathbb{N}}$ of real numbers, and Følner sequence $(\Phi_N)_{N\in\mathbb{N}}$ in $\mathbb{Z}$ we have

$$\lim_{N\to\infty}\sup_{p\in\mathbb{R}_{k-1}[t]}\Big\|\mathbb{E}_{n\in\Phi_N}c_{N,n}\,T^{[n^k\alpha+p(n)]}f_1\cdot T^{2[n^k\alpha+p(n)]}f_2\cdot\ldots\cdot T^{\ell[n^k\alpha+p(n)]}f_\ell\Big\|_{L^2(\mu)}=0.$$

The main step in the deduction of Theorem 2.4 from Proposition 4.1 is carried out in Lemma 4.4 below. Before delving into the proof of this lemma we mention two useful ingredients that will be used in its proof. The first one was proved in [14]:

Lemma 4.2. *Let $a\in\mathcal{H}$ have polynomial growth and satisfy $|a(t)-p(t)|\succ\log t$ for every $p\in\mathbb{R}[t]$.*

Then for some $k\in\mathbb{N}$ we have

$$|a^{(k+1)}(t)| \text{ decreases to } 0,\quad 1/t^k\prec a^{(k)}(t)\prec 1,\quad \text{and}\quad (a^{(k+1)}(t))^k\prec(a^{(k)}(t))^{k+1}. \tag{26}$$

The second is the following simple result:

Lemma 4.3. *Let $(V(n))_{n\in\mathbb{N}}$ be a bounded sequence of vectors on a normed space. Suppose that*

$$\lim_{N\to\infty}\big(\mathbb{E}_{N\leq n\leq N+l(N)}V(n)\big)=0$$

for some positive function $l(t)$ with $l(t)\prec t$. Then

$$\lim_{N\to\infty}\mathbb{E}_{1\leq n\leq N}V(n)=0.$$

Proof. We can cover the interval $[1,N]$ by a union of non-overlapping intervals of the form $[k,k+l(k)]$, we denote this union by I_N. Since $l(t)\prec t$ and the sequence $(V(n))_{n\in\mathbb{N}}$ is bounded we have that

$$\lim_{N\to\infty}\mathbb{E}_{1\leq n\leq N}V(n)=\lim_{N\to\infty}\mathbb{E}_{n\in I_N}V(n).$$

Using our assumption, one easily gets that the limit $\lim_{N\to\infty}\mathbb{E}_{n\in I_N}V(n)$ is 0, finishing the proof. □

Lemma 4.4. *Let $(V(n))_{n\in\mathbb{N}}$ be a bounded sequence of vectors on a normed space. Suppose that for every $k\in\mathbb{N}$ and bounded two parameter sequence $(c_{N,n})_{N,n\in\mathbb{N}}$ of real numbers we have*

$$\lim_{N\to\infty}\sup_{p\in\mathbb{R}_{k-1}[t],}\left\|\mathbb{E}_{1\leq n\leq N}c_{N,n}\,V(n^k+[p(n)])\right\|=0.$$

Then if $a\in\mathcal{H}$ has polynomial growth and satisfies $|a(t)-p(t)|\succ\log t$ for every $p\in\mathbb{R}[t]$, and $(c_n)_{n\in\mathbb{N}}$ is any bounded sequence of real numbers, we have

$$\lim_{N\to\infty}\mathbb{E}_{1\leq n\leq N}c_n\,V([a(n)])=0.$$

Proof. For convenience we assume that $c_n=1$ for every $n\in\mathbb{N}$, the proof is similar in the general case. By Lemma 4.3 it suffices to show that

$$\lim_{N\to\infty}\mathbb{E}_{N<n\leq N+l(N)}V([a(n)])=0 \tag{27}$$

for some function $l(t)$ with $l(t)\prec t$ (we shall impose more conditions on $l(t)$ as the argument proceeds).

Let $k\in\mathbb{N}$ be such that the conclusion of Lemma 4.2 is satisfied, namely,

$$|a^{(k+1)}(t)|\text{ decreases to }0,\quad 1/t^k\prec a^{(k)}(t)\prec 1,\quad\text{ and }\quad (a^{(k+1)}(t))^k\prec(a^{(k)}(t))^{k+1}. \tag{28}$$

For convenience we are going to assume that $a^{(k)}(t)$ is eventually positive, the proof is similar in the other case.

Using the Taylor expansion of $a(t)$ around the point $t=N$ we get for every $n\in\mathbb{N}$ that

$$a(N+n)=a(N)+na'(N)+\cdots+\frac{n^k}{k!}a^{(k)}(N)+\frac{n^{k+1}}{(k+1)!}a^{(k+1)}(\xi_n) \tag{29}$$

for some $\xi_n\in[N,N+n]$. Since $|a^{(k+1)}(t)|$ is eventually decreasing, for every large N we have $|a^{(k+1)}(\xi_n)|\leq|a^{(k+1)}(N)|$. It follows that if $l(t)$ satisfies

$$(l(t))^{k+1}a^{(k+1)}(t)\prec 1, \tag{30}$$

then for every large N and $n=1,\ldots,[l(N)]$ we have

$$[a(N+n)]=\Big[a(N)+na'(N)+\cdots+\frac{n^k}{k!}a^{(k)}(N)\Big]+e_N(n) \tag{31}$$

where the error terms $e_N(n)$ take values in the set $\{0,-1\}$ (we used that $a^{(k+1)}(t)$ is eventually negative). For $t\in[0,1]$ and x positive we have

$$\left|\frac{1}{(x+t)^k}-\frac{1}{x^k}\right|\leq\frac{k}{x^{k+1}},$$

therefore if

$$d(t)=\frac{k!}{a^{(k)}(t)},$$

then setting $x = [\sqrt[k]{d(N)}]$ and $t = \{\sqrt[k]{d(N)}\}$ in the previous estimate we get

$$\left|\frac{a^{(k)}(N)}{k!} - \frac{1}{[\sqrt[k]{d(N)}]^k}\right| \leq \frac{k}{[\sqrt[k]{d(N)}]^{k+1}} \sim (a^{(k)}(N))^{1+\frac{1}{k}}.$$

From this estimate and (31), it follows that if

$$(l(t))^k (a^{(k)}(t))^{1+\frac{1}{k}} \prec 1, \tag{32}$$

then for every large N and $n = 1, \dots, [l(N)]$ we have

$$[a(N+n)] = \Big[a(N) + na'(N) + \cdots + \frac{n^k}{[\sqrt[k]{d(N)}]^k}\Big] + \tilde{e}_N(n)$$

where the error terms $\tilde{e}_N(n)$ take values in the set $\{-2, -1, 0, 1\}$. Hence, in order to establish (27) it suffices to show that

$$\lim_{N\to\infty} \sup_{p\in\mathbb{R}_{k-1}[t]} \left\|\mathbb{E}_{1\leq n\leq l(N)} V\Big(\Big[\frac{n^k}{[\sqrt[k]{d(N)}]^k} + p(n)\Big] + \tilde{e}_N(n)\Big)\right\| = 0. \tag{33}$$

Next notice that (33) follows if we show that for every bounded sequence $(c_{N,n})_{N,n\in\mathbb{N}}$ we have

$$\lim_{N\to\infty} \sup_{p\in\mathbb{R}_{k-1}[t]} \left\|\mathbb{E}_{1\leq n\leq l(N)} c_{N,n}\, V\Big(\Big[\frac{n^k}{[\sqrt[k]{d(N)}]^k} + p(n)\Big]\Big)\right\| = 0. \tag{34}$$

Indeed, it suffices to use (34) when $c_{N,n} = \mathbf{1}_{\{k\colon \tilde{e}_N(k)=i\}}(n)$ for $i = -2, -1, 0, 1$, and then add the corresponding identities.

We perform one last maneuver by rewriting (34) in a more convenient form. Notice that every integer between 1 and $l(N)$ can be represented as $[\sqrt[k]{d(N)}]n + r$ with $0 \leq n \leq \tilde{l}(N) = l(N)/[\sqrt[k]{d(N)}]$ and $1 \leq r \leq [\sqrt[k]{d(N)}]$. Furthermore, if we choose $l(t)$ so that

$$(l(t))^k a^{(k)}(t) \succ 1, \tag{35}$$

then we have $\tilde{l}(N) \to \infty$. Since for every bounded sequence of vectors $V(n)$ the average $\mathbb{E}_{1\leq n\leq b(N)} V(n)$ is equal to $\mathbb{E}_{1\leq r\leq r(N)} \mathbb{E}_{1\leq n\leq \tilde{l}(N)} V([\sqrt[k]{d(N)}]n + r)$ (up to negligible terms), an easy computation (similar to the one used in the model example) shows that in order to establish (34) it suffices to show that

$$\lim_{N\to\infty} \sup_{p\in\mathbb{R}_{k-1}[t],} \left\|\mathbb{E}_{1\leq n\leq N} c_{N,n}\, V([n^k + p(n)])\right\|_{L^2(\mu)} = 0. \tag{36}$$

Summarizing, we have reduced matters to establishing (36), which holds by our assumption, provided that there exists a function $l(t)$ that satisfies all the conditions imposed previously, namely, $l(t) \prec t$ and the conditions stated in equations (30), (32), and (35). Equivalently, the function $l(t)$ must satisfy

$$\frac{1}{(a^{(k)}(t))^{\frac{1}{k}}} \prec l(t), \quad l(t) \prec t, \quad l(t) \prec \frac{1}{(a^{(k+1)}(t))^{\frac{1}{k+1}}}, \quad \text{and} \quad l(t) \prec \frac{1}{(a^{(k)}(t))^{\frac{1}{k}+\frac{1}{k^2}}}.$$

That such a function $l(t)$ exists follows from the second and third conditions in (28), thus completing the proof. $\square$

With the help of Proposition 4.1 and Lemma 4.4 it is now easy to prove Theorem 2.4.

Proof of Theorem 2.4. Let $(X, \mathcal{X}, \mu, T)$ be a system and suppose that at least one of the functions $f_1, \ldots, f_\ell \in L^\infty(\mu)$ is orthogonal to the nilfactor $\mathcal{Z}$. Let $a \in \mathcal{H}$ have polynomial growth and satisfy $a(t) \succ \log t$. We have to show that

$$\lim_{N\to\infty} \mathbb{E}_{1\leq n\leq N} T^{[a(n)]} f_1 \cdot \ldots \cdot T^{\ell[a(n)]} f_\ell = 0 \tag{37}$$

where the convergence takes place in $L^2(\mu)$.

Combining Proposition 4.1 and Lemma 4.4 we immediately get that (37) holds if $a \in \mathcal{H}$ has polynomial growth and satisfies $|a(t) - p(t)| \succ \log t$ for every $p \in \mathbb{R}[t]$. Therefore, it remains to deal with the case where $a(t) = p(t) + e(t)$, where $p \in \mathbb{R}[t]$ is non-constant and $e(t) \ll \log t$.

Suppose first that $e(t)$ is bounded. Then $e(n) \to c$ for some $c \in \mathbb{R}$, and as a result for every large $n \in \mathbb{N}$ we have $[a(n)] = [p(n) + c] + \tilde{e}(n)$ for some sequence $(\tilde{e}(n))_{n\in\mathbb{N}}$ with $\tilde{e}(n) \in \{0, \pm 1\}$. Using this, we deduce that (37) follows from Proposition 4.1.

The last case to consider is when $1 \prec e(t) \prec \log t$. Let $I_m = \{n \in \mathbb{N} \colon [e(n)] = m\}$. Since $e(n+1) - e(n) \to 0$ (this follows from $e'(t) \to 0$ and the mean value theorem), and $e(n) \to \infty$, it follows that for every large m the set I_m is an integer interval with length that increases to infinity. Notice also that for $n \in I_m$ we have $[a(n)] = [p(n)] + m + \tilde{e}(n)$ for some sequence $(\tilde{e}(n))_{n\in\mathbb{N}}$ with $\tilde{e}(n) \in \{0, \pm 1\}$. Using this, we deduce that (37) follows from Proposition 4.1. This completes the proof. $\square$

4.3. **Proof of the polynomial ergodic theorem.** Let $\mathcal{P} = \{p_1, \ldots, p_\ell\}$ be a family of polynomials with real coefficients. We say that the family $\mathcal{P}$ consists of *non-constant and essentially distinct* polynomials, if all the polynomials and their pairwise differences have positive degree. The maximum degree of the polynomials is called the *degree* of the polynomial family, and is denoted by $\deg(\mathcal{P})$. Given a polynomial family $\mathcal{P}$, let $\mathcal{P}_i$ be the subfamily of polynomials of degree i in $\mathcal{P}$. We let w_i denote the number of distinct leading coefficients that appear in the family $\mathcal{P}_i$. The vector $(d, w_d, \ldots, w_1)$, where $d = \deg(\mathcal{P})$, is called the *type* of the polynomial family $\mathcal{P}$.

We will use an induction scheme, often called PET induction (Polynomial Exhaustion Technique), on types of polynomial families that was introduced by Bergelson in [3]. We order the set of all possible types lexicographically, meaning, $(d, w_d, \ldots, w_1) > (d', w'_d, \ldots, w'_1)$ if and only if in the first instance where the two vectors disagree the coordinate of the first vector is greater than the coordinate of the second vector.

Given a family of non-constant essentially distinct polynomials $\mathcal{P} = \{p_1, \ldots p_\ell\}$, a positive integer h, and $p \in \mathcal{P}$, we form a new family $\mathcal{P}(p, h)$ as follows: We start with the family of polynomials

$$\{p_1(t+h) - p(t), \ldots, p_\ell(t+h) - p(t), p_1(t) - p(t), \ldots, p_\ell(t) - p(t)\},$$

and successively remove the smallest number of polynomials so that the resulting family consists of non-constant, essentially distinct polynomials. Then for every large h the function $p_i(t+h) - p(t)$ will be removed if and only if p_i is linear (then $(p_i(t+h) - p(t)) - (p_i(t) - p(t)) = p_i(h)$), and the function $p_i(t) - p(t)$ will be removed if and only if $p = p_i$.

Example 1. If $\mathcal{P} = \{t, 2t, t^2\}$ and $p(t) = t$, then we start with the family of polynomials

$$\{h, t + 2h, (t+h)^2 - t, 0, t, t^2 - t\}$$

and remove the first, second, and fourth polynomials to get

$$\mathcal{P}(t, h) = \{(t+h)^2 - t, t, t^2 - t\}.$$

Notice that the family $\mathcal{P}$ has type $(2,1,2)$, and the family $\mathcal{P}(t,h)$ has smaller type, namely, $(2,1,1)$.

The main step in the proof of Proposition 4.1 is carried out in Lemma 4.7. This lemma is proved using induction on the type of the family of functions involved. In order to carry out the inductive step we will use the following:

Lemma 4.5. *Let $\mathcal{P}=\{p_1,\dots p_\ell\}$ be family of non-constant essentially distinct polynomials, and suppose that $deg(p_1)=deg(\mathcal{P})\geq 2$.*

Then there exists $p\in\mathcal{P}$ such that for every large h the family $\mathcal{P}(p,h)$ has type smaller than that of $\mathcal{P}$, and $deg(p_1(t+h)-p(t))=deg(\mathcal{P}(p,h))$.

Remark. Since $\deg(p_1)\geq 2$, no-matter what the choice of p will be, the polynomial $p_1(t+h)-p(t)$ is going to be an element of the family $\mathcal{P}(p,h)$ for every large h.

Proof. Suppose first that $\deg(p_i)<\deg(p_1)$ for some $i\in\{2,\dots,\ell\}$. Let i_0 be such that the polynomial p_{i_0} has minimal degree. Then $p=p_{i_0}$ has the advertised property.

Otherwise, all the polynomials have the same degree, in which case for $i=2,\dots,\ell$ we have $p_i(t)=\alpha_i p_1(t)+q_i(t)$ for some non-zero real numbers $\alpha_2,\dots,\alpha_\ell$ and polynomials q_i with $\deg(q_i)<\deg(p_1)$. If $\alpha_{i_0}\neq 1$ for some $i_0\in\{2,\dots,\ell\}$, then $p=p_{i_0}$ has the advertised property. If $\alpha_i=1$ for $i=2,\dots,\ell$, let i_0 be such that the function q_{i_0} has maximal degree. Then $p=p_{i_0}$ has the advertised property. This completes the proof. □

We are also going to use a variation of the classical elementary lemma of van der Corput. Its proof is a straightforward modification of the one given in [3], therefore we omit it.

Lemma 4.6. *Let $\{v_{N,n}\}_{N,n\in\mathbb{N}}$ be a bounded sequence of vectors in a Hilbert space, and $(\Phi_N)_{N\in\mathbb{N}}$ be a Følner sequence of subsets of $\mathbb{N}$. For every $h\in\mathbb{N}$ we set*

$$b_h=\overline{\lim}_{N\to\infty}\Big|\mathbb{E}_{n\in\Phi_N}<v_{N,n+h},v_{N,n}>\Big|.$$

Then

$$\overline{\lim}_{N\to\infty}\left\|\mathbb{E}_{n\in\Phi_N}v_{N,n}\right\|^2\leq 4\,\overline{\lim}_{H\to\infty}\mathbb{E}_{1\leq h\leq H}b_h.$$

To state our next result it will be convenient to introduce some notation. For $N\in\mathbb{N}$ let $\mathcal{P}_N=\{p_{1,N},\dots,p_{\ell,N}\}$ be a family of polynomials with real coefficients. We say that the collection $(\mathcal{P}_N)_{N\in\mathbb{N}}$ is *"nice"* if for every $N\in\mathbb{N}$ the polynomials $p_{i,N}$ and $p_{i,N}-p_{j,N}$ (for $i\neq j$) are non-constant and their leading coefficients are independent of N.

Lemma 4.7. *Let $(\{p_{1,N},\dots,p_{\ell,N}\})_{N\in\mathbb{N}}$ be a "nice" collection of polynomial families. Let $(X,\mathcal{X},\mu,T)$ be a system, and suppose that one of the functions $f_1,\dots,f_\ell\in L^\infty(\mu)$ is orthogonal to the nilfactor $\mathcal{Z}$.*

Then for every Følner sequence $(\Phi_N)_{N\in\mathbb{N}}$ in $\mathbb{Z}$ and bounded sequence $(c_{N,n})_{N,n\in\mathbb{N}}$ we have

$$\lim_{N\to\infty}\mathbb{E}_{n\in\Phi_N}c_{N,n}\,T^{[p_{1,N}(n)]}f_1\cdot\ldots\cdot T^{[p_{\ell,N}(n)]}f_\ell=0 \tag{38}$$

where the convergence takes place in $L^2(\mu)$.

Remark. In the special case where $p_{i,N}=p_i$ for $i=1,\dots,\ell$ we get a different proof[9] of the known result that the nilfactor $\mathcal{Z}$ is characteristic for any family of non-constant, essentially distinct polynomials of a single variable.

[9]In contrast with the proof given in [29], we do not have to work with polynomials of several variables (which was a key trick in [29]) in order to prove the single variable result.

Proof. Without loss of generality we can assume that the function f_1 is orthogonal to the nilfactor $\mathcal{Z}$. Furthermore, we can assume that $\|f_i\|_\infty \leq 1$ for $i = 1, \ldots, \ell$, and $|c_{N,n}| \leq 1$ for every $N, n \in \mathbb{N}$. It will be crucial for our argument to assume that the polynomial $p_{1,N}$ has maximal degree within the family $\mathcal{P}_N = \{p_{1,N}, \ldots, p_{\ell,N}\}$. To get this extra assumption it is convenient to somewhat modify our goal; instead of (38) we shall prove that

$$\lim_{N\to\infty} \sup_{\|f_0\|_\infty, \|f_2\|_\infty, \ldots, \|f_\ell\|_\infty \leq 1} \mathbb{E}_{n\in\Phi_N} \Big| \int f_0 \cdot T^{[p_{1,N}(n)]} f_1 \cdot T^{[p_{2,N}(n)]} f_2 \cdot \ldots \cdot T^{[p_{\ell,N}(n)]} f_\ell \, d\mu \Big| = 0. \tag{39}$$

Notice first that (39) implies (38). Indeed, (39) gives that

$$\lim_{N\to\infty} \mathbb{E}_{n\in\Phi_N} c_{N,n} \int f_{0,N} \cdot T^{[p_{1,N}(n)]} f_1 \cdot T^{[p_{2,N}(n)]} f_2 \cdot \ldots \cdot T^{[p_{\ell,N}(n)]} f_\ell \, d\mu = 0 \tag{40}$$

whenever $f_{0,N} \in L^\infty(\mu)$ satisfies $\|f_{0,N}\|_\infty \leq 1$ for $N \in \mathbb{N}$. Using (40) with the conjugate of the function $\mathbb{E}_{n\in\Phi_N} c_{N,n}\, T^{[p_{1,N}(n)]} f_1 \cdot T^{[p_{2,N}(n)]} f_2 \cdot \ldots \cdot T^{[p_{\ell,N}(n)]} f_\ell$ in place of the function $f_{0,N}$ (for every $N \in \mathbb{N}$), we get (38).

Next we claim that when proving (39) we can assume that the polynomial $p_{1,N}$ has maximal degree within the family $\mathcal{P}_N$. Indeed, if this is not the case, then $\deg(p_{1,N}) < \deg(p_{i,N})$ for some $i = 2, \ldots, \ell$, say this happens for $i = \ell$. After factoring out the transformation $T^{[p_{\ell,N}(n)]}$ we see that (39) can be rewritten as

$$\begin{aligned} \lim_{N\to\infty} \sup_{\|f_0\|_\infty, \|f_2\|_\infty, \ldots, \|f_\ell\|_\infty \leq 1} \mathbb{E}_{n\in\Phi_N} \Big| \int f_\ell \cdot T^{[-p_{\ell,N}(n)]+e_{0,N}(n)} f_0 \cdot T^{[p_{1,N}(n)-p_{\ell,N}(n)]+e_{1,N}(n)} f_1 \cdot \\ T^{[p_{2,N}(n)-p_{\ell,N}(n)]+e_{2,N}(n)} f_2 \cdot \ldots \cdot T^{[p_{\ell-1,N}(n)-p_{\ell,N}(n)]+e_{\ell-1,N}(n)} f_{\ell-1} \, d\mu \Big| = 0 \end{aligned} \tag{41}$$

for some error terms $e_{i,N}(n)$ with values in the set $\{0, 1\}$. Furthermore, notice that (41) follows if we show that

$$\begin{aligned} \lim_{N\to\infty} \sup_{\|f_0\|_\infty, \|f_2\|_\infty, \ldots, \|f_\ell\|_\infty \leq 1} \mathbb{E}_{n\in\Phi_N} \Big| \int f_\ell \cdot T^{[-p_{\ell,N}(n)]} f_0 \cdot T^{[p_{1,N}(n)-p_{\ell,N}(n)]} f_1 \cdot \\ T^{[p_{2,N}(n)-p_{\ell,N}(n)]} f_2 \cdot \ldots \cdot T^{[p_{\ell-1,N}(n)-p_{\ell,N}(n)]} f_{\ell-1} \, d\mu \Big| = 0. \end{aligned} \tag{42}$$

Since the collection of polynomial families $(\mathcal{P}'_N)_{N\in\mathbb{N}}$, where

$$\mathcal{P}'_N = \{-p_{\ell,N}, p_{1,N} - p_{\ell,N}, \ldots, p_{\ell-1,N} - p_{\ell,N}\},$$

is also "nice", and the polynomial $p_{1,N} - p_{\ell,N}$ has maximal degree within the family $\mathcal{P}'_N$, the claim follows.

Summarizing, we have reduced matters to establishing (39) for every system, assuming that the function $f_1 \in L^\infty(\mu)$ is orthogonal to the nilfactor $\mathcal{Z}$ and the polynomial $p_{1,N}$ has maximal degree within the family $\mathcal{P}_N$. We shall do this by using induction on the type of the family of polynomials $\mathcal{P}_N$ (the type of this family is independent of N).

The case where all the polynomials have degree 1 can be treated as in the proof of (52) in Proposition 5.3 below. (For linear functions, the same argument works for any Følner sequence Φ_N in place of the intervals $[1, N]$; we leave the routine details to the reader.)

Now let $d \geq 2$ and suppose that the statement holds for every "nice" collection of polynomial families with type smaller than $(d, w_d, \ldots, w_1)$. Let $(\mathcal{P}_N)_{N\in\mathbb{N}}$, where $\mathcal{P}_N = \{p_{1,N}, \ldots, p_{\ell,N}\}$, be a "nice" collection of polynomial families with type $(d, w_d, \ldots, w_1)$.

Using the Cauchy-Schwarz inequality we see that (39) follows if we show that

$$\lim_{N\to\infty} \sup_{\|f_0\|_\infty, \|f_2\|_\infty,\ldots,\|f_\ell\|_\infty\leq 1} \mathbb{E}_{n\in\Phi_N}\Big|\int f_0\cdot T^{[p_{1,N}(n)]}f_1\cdot T^{[p_{2,N}(n)]}f_2\cdot\ldots\cdot T^{[p_{\ell,N}(n)]}f_\ell\, d\mu\Big|^2=0.$$

This last identity can be rewritten as

$$\lim_{N\to\infty} \sup_{\|f_0\|_\infty, \|f_2\|_\infty,\ldots,\|f_\ell\|_\infty\leq 1} \mathbb{E}_{n\in\Phi_N}\int F_0\cdot S^{[p_{1,N}(n)]}F_1\cdot S^{[p_{2,N}(n)]}F_2\cdot\ldots\cdot S^{[p_{\ell,N}(n)]}F_\ell\, d(\mu\times\mu)=0$$

where $S=T\times T$ and $F_i=f_i\otimes\overline{f}_i$ for $i=0,1,\ldots,\ell$. Using the Cauchy-Schwarz inequality we see that it suffices to show that

$$\lim_{N\to\infty} \sup_{\|F_2\|_\infty,\ldots,\|F_\ell\|_\infty\leq 1} \Big\|\mathbb{E}_{n\in\Phi_N}\cdot S^{[p_{1,N}(n)]}F_1\cdot S^{[p_{2,N}(n)]}F_2\cdot\ldots\cdot S^{[p_{\ell,N}(n)]}F_\ell\Big\|_{L^2(\mu\times\mu)}=0 \tag{43}$$

where $F_1=f_1\otimes\overline{f}_1$. We choose functions $F_{i,N}$, $i=2,\ldots,\ell$, with sup norm at most 1, so that the value of the norms in (43) is $1/N$ close to the supremum and use Lemma 4.6. We get that (43) follows if we show that for every large h we have

$$\lim_{N\to\infty} \sup_{\|F_2\|_\infty,\ldots,\|F_\ell\|_\infty\leq 1} \Big|\mathbb{E}_{n\in\Phi_N}\int S^{[p_{1,N}(n+h)]}F_1\cdot S^{[p_{2,N}(n+h)]}F_2\cdot\ldots\cdot S^{[p_{\ell,N}(n+h)]}F_\ell\cdot$$
$$S^{[p_{1,N}(n)]}\overline{F}_1\cdot S^{[p_{2,N}(n)]}\overline{F}_2\cdot\ldots\cdot S^{[p_{\ell,N}(n)]}\overline{F}_\ell\, d(\mu\times\mu)\Big|=0.$$

Factoring out the transformation $S^{[p_N(n)]}$, where $p_N=p_{i,N}$ for some $i\in\{1,\ldots,\ell\}$ is chosen as in Lemma 4.5 (the choice of i is independent of N), we can rewrite the last identity as

$$\lim_{N\to\infty} \sup_{\|F_2\|_\infty,\ldots,\|F_\ell\|_\infty\leq 1} \mathbb{E}_{n\in\Phi_N}\Big|\int S^{[p_{1,N}(n+h)-p_N(n)]+e_{1,N}(h,n)}F_1\cdot\ldots\cdot S^{[p_{\ell,N}(n+h)-p_N(n)]+e_{\ell,N}(h,n)}F_\ell\cdot$$
$$S^{[p_{1,N}(n)-p_N(n)]+e_{\ell+1,N}(h,n)}\overline{F}_1\cdot\ldots\cdot S^{[p_{\ell,N}(n)-p_N(n)]+e_{2\ell,N}(h,n)}\overline{F}_\ell\, d(\mu\times\mu)\Big|=0$$

for some error terms $e_{i,N}(h,n)$ with values in $\{0,1\}$. For every fixed $h,N\in\mathbb{N}$ we can partition the integers into a finite number of sets, that depend only on ℓ, where all sequences $e_{i,N}(h,n)$ are constant. Therefore, the last identity follows if we show that for every large h we have

$$\lim_{N\to\infty} \sup_{\|F_2\|_\infty,\ldots,\|F_\ell\|_\infty\leq 1} \mathbb{E}_{n\in\Phi_N}\Big|\int S^{[p_{1,N}(n+h)-p_N(n)]}F_1\cdot\ldots\cdot S^{[p_{\ell,N}(n+h)-p_N(n)]}F_\ell\cdot \tag{44}$$
$$S^{[p_{1,N}(n)-p_N(n)]}F_1\cdot\ldots\cdot S^{[p_{\ell,N}(n)-p_N(n)]}\overline{F}_\ell\, d(\mu\times\mu)\Big|=0.$$

Next notice that if the polynomial $p_{i,N}$ have degree 1 (this can only happen for $i\neq 1$), then $p_{i,N}(n+h)=p_{i,N}(n)+c_N(h)$ for some $c_N(h)\in\mathbb{R}$. Hence, for those values of i we can write

$$S^{[p_{i,N}(n+h)-p_N(n)]}F_i\cdot S^{[p_{i,N}(n)-p_N(n)]}\overline{F}_i=S^{[p_{i,N}(n)-p_N(n)]}(S^{[c_N(h)]+e_N(h,n)}F_i\cdot\overline{F}_i),$$

for some error terms $e_N(h,n)\in\{0,1\}$. As explained before, because the error terms $e_N(h,n)$ take values in a finite set, when proving (44) we can assume that $e_N(h,n)=0$. Therefore, we have reduced matters to showing that for every large h we have

$$\lim_{N\to\infty} \sup_{\|F_0\|,\|F_2\|_\infty,\ldots,\|F_r\|_\infty\leq 1} \Big|\mathbb{E}_{n\in\Phi_N}\int F_0\cdot S^{[p_{1,N}(n+h)-p_N(n)]}F_1\cdot S^{[q_{2,N}(n)]}F_2\cdot\ldots\cdot \tag{45}$$
$$S^{[q_{r,N}(n)]}F_r\, d(\mu\times\mu)\Big|=0.$$

for some $r \in \mathbb{N}$, where all the polynomials involved are elements of the collection of polynomial families $(\mathcal{P}_N(p_N, h))_{N\in\mathbb{N}}$ (defined on the beginning of this subsection). For every large h this new collection of polynomial families is "nice" and by Lemma 4.5 it has type smaller than that of $(\mathcal{P}_N)_{N\in\mathbb{N}}$. Furthermore, the degree of the polynomial $p_{1,N}(n+h)-p_N(n)$ is maximal within the family $\mathcal{P}_N(p_N, h)$. Since f_1 is orthogonal to the factor $\mathcal{Z}(T)$, we have that $F_1 = f_1 \otimes \overline{f}_1$ is orthogonal to the factor $\mathcal{Z}(S)$. Therefore, the induction hypothesis applies, and verifies (45) for every large h. This completes the induction and the proof. □

We are now ready to give the proof of Proposition 4.1.

Proof of Proposition 4.1. Let $(X, \mathcal{X}, \mu, T)$ be a system, and suppose that at least one of the functions $f_1, \dots, f_\ell \in L^\infty(\mu)$ is orthogonal to the nilfactor $\mathcal{Z}$. Our goal is to show that for every bounded two variable sequence $(c_{N,n})_{n\in\mathbb{N}}$ we have

$$\lim_{N\to\infty} \sup_{p\in\mathbb{R}_{k-1}[t]} \Big\|\mathbb{E}_{n\in\Phi_N} c_{N,n}\, T^{[n^k\alpha+p(n)]} f_1 \cdot T^{2[n^k\alpha+p(n)]} f_2 \cdot \ldots \cdot T^{\ell[n^k\alpha+p(n)]} f_\ell\Big\|_{L^2(\mu)} = 0. \tag{46}$$

We choose polynomials $p_N \in \mathbb{R}_{k-1}[t]$, so that the norm in (46) is $1/N$ close to the supremum. Then (46) takes the form

$$\lim_{N\to\infty} \Big\|\mathbb{E}_{n\in\Phi_N} c_{N,n}\, T^{[n^k\alpha+p_N(n)]} f_1 \cdot T^{2[n^k\alpha+p_N(n)]} f_2 \cdot \ldots \cdot T^{\ell[n^k\alpha+p_N(n)]} f_\ell\Big\|_{L^2(\mu)} = 0. \tag{47}$$

Equation (47) can be rewritten as

$$\lim_{N\to\infty} \mathbb{E}_{n\in\Phi_N} c_{N,n} T^{[n^k\alpha+p_N(n)]} f_1 \cdot T^{[2(n^k\alpha+p_N(n))]+e_{2,N}(n)} f_2 \cdot \ldots \cdot T^{[\ell(n^k\alpha+p_N(n))]+e_{\ell,N}(n)} f_\ell = 0 \tag{48}$$

where convergence takes place in $L^2(\mu)$ and the error terms $e_{2,N}(n), \dots, e_{\ell,N}(n)$ take values in the set $\{0, -1, \dots, -\ell\}$. Using Lemma 4.7 we deduce that (48) holds, completing the proof. □

5. Characteristic factors for several sequences

A crucial step in the proof of Theorem 2.6 is to show that the nilfactor $\mathcal{Z}$ is characteristic for the related multiple ergodic averages. This is the context of Theorem 2.9 which we are going to prove in this section. Our proof extends an argument used in [7] where a similar result was verified for weakly mixing systems. Since there are a few non-trivial extra complications in our case, we shall give our proof in more or less full detail, referring the reader to [7] only when an argument we need is a straightforward modification of one used there.

5.1. The sub-linear case.

First, we are going to prove Theorem 2.9 in the case where all the functions have at most linear growth.

We first give two lemmas that will be used in our proof. The first is implicit in [7]:

Lemma 5.1. *Let $(V_N(n))_{N,n\in\mathbb{N}}$ be a bounded two parameter sequence of vectors in a normed space and suppose that $a \in \mathcal{H}$ satisfies $t^\varepsilon \prec a(t) \prec t$ for some $\varepsilon > 0$.*

Then

$$\lim_{N\to\infty} \Big(\mathbb{E}_{1\le n\le N} V_N([a(n)]) - \mathbb{E}_{1\le n\le N} V_N(n)\Big) = 0.$$

Proof. Without loss of generality we can assume that the limits of both averages exist.

Let $w(n) = |m \in \mathbb{N}\colon [a(m)] = n|$ and $W(n) = \sum_{m=1}^n w(m)$. Using our hypothesis on $a(t)$ it is not hard to verify (for the details see Lemma 2.5 in [7]) that

$$\frac{w(n)}{W(n)} \to 0 \quad \text{and} \quad \frac{nw(n)}{W(n)} \text{ is bounded.} \tag{49}$$

Using the definition of $w(n)$ and $W(n)$, that $w(n)/W(n) \to 0$, and that $(V_N(n))_{N,n\in\mathbb{N}}$ is bounded, we conclude that

$$\lim_{N\to\infty} \mathbb{E}_{1\leq n\leq N} V_N([a(n)]) = \lim_{N\to\infty} \frac{1}{W(N)} \sum_{n=1}^{N} w(n)V_N(n).$$

Let $A(N) = \mathbb{E}_{n=1}^{N} V_N(n)$ and $A = \lim_{N\to\infty} A(N)$. Using summation by parts we have

$$\frac{1}{W(N)} \sum_{n=1}^{N} w(n)V_N(n) = \frac{1}{W(N)}\Big(Nw(N)A(N) + \sum_{n=2}^{N} n(w(n-1) - w(n))A(n)\Big).$$

Letting $V_N(n) = 1$ in the previous identity we find that

$$1 = \frac{1}{W(N)}\Big(Nw(N) + \sum_{n=2}^{N} n(w(n-1) - w(n))\Big).$$

It follows that

$$\Big|\frac{1}{W(N)} \sum_{n=1}^{N} w(n)V_N(n) - A\Big| =$$

$$\frac{1}{W(N)}\Big|Nw(N)(A(N) - A) + \sum_{n=2}^{N} n(w(n-1) - w(n))(A(n) - A)\Big| \leq$$

$$\underbrace{\frac{Nw(N)}{W(N)}|(A(N) - A)|}_{I_1(N)} + \underbrace{\frac{1}{W(N)}\Big|\sum_{n=2}^{N} n(w(n-1) - w(n))(A(n) - A)\Big|}_{I_2(N)}.$$

Since $Nw(N)/W(N)$ is bounded and $A(N) \to A$ we have that $I_1(N) \to 0$. Furthermore, using that $A(N) \to A$, and $W(N) \to \infty$, we see that for given $\varepsilon > 0$ we have

$$I_2(N) \leq \frac{N\sum_{n=2}^{N} |w(n-1) - w(n)|}{W(N)}\,\varepsilon + o_N(1) = \frac{N(w(N) - w(1))}{W(N)}\,\varepsilon + o_N(1)$$

where the last equality holds because $w(n)$ is increasing. By (49), the last fraction is bounded by $C\varepsilon$ for some positive $C \in \mathbb{R}$. This proves that

$$\lim_{N\to\infty} \Big|\frac{1}{W(N)} \sum_{n=1}^{N} w(n)V(n) - A\Big| \leq C\varepsilon.$$

Since the constant C depends only on the function $a(t)$, and ε is arbitrary, the proof is complete. $\square$

The second lemma is a well known estimate, we prove it for completeness.

Lemma 5.2. *Let $(X, \mathcal{X}, \mu, T)$ be a system and $f_1 \in L^\infty(\mu)$.*
Then we have

$$\limsup_{N\to\infty} \sup_{\|f_0\|_\infty\leq 1} \mathbb{E}_{1\leq n\leq N}\Big|\int f_0 \cdot T^n f_1 \; d\mu\Big| \leq \|f_1\|_2. \tag{50}$$

Proof. Using the Cauchy-Schwarz inequality we see that
$$\Big(\mathbb{E}_{1\le n\le N}\Big|\int f_0\cdot T^n f_1\, d\mu\Big|\Big)^2 \le \mathbb{E}_{1\le n\le N}\Big|\int f_0\cdot T^n f_1\, d\mu\Big|^2.$$
The last average can be rewritten as
$$\mathbb{E}_{1\le n\le N}\int F_0\cdot S^n F_1\, d(\mu\times\mu)$$
where $S=T\times T$, $F_0=f_0\otimes\overline{f}_0$, and $F_1=f_1\otimes\overline{f}_1$. Assuming that $\|f_0\|_\infty\le 1$, and using the Cauchy-Schwarz inequality again, we find that the last average is bounded by
$$\|\mathbb{E}_{1\le n\le N}S^nF_1\|_{L^2(\mu\times\mu)}\,.$$
Taking limits, and using the ergodic theorem, we conclude that the square of the limit in (50) is bounded by
$$\big\|\mathbb{E}_{\mu\times\mu}(f_1\otimes\overline{f}_1|\mathcal{I}(S))\big\|_{L^2(\mu\times\mu)}=|\!|\!|f_1|\!|\!|_2^2$$
where the last identity follows from the second identity in (21) and the definition of $|\!|\!|f|\!|\!|_1$. This establishes the advertised estimate. □

In the proof of Proposition 5.3 we are going to use the symbol $\ll_{w_1,\ldots,w_k}$ when some expression is majorized by some other expression and the implied constant depends on the parameters $w_1,\ldots,w_k$.

Proposition 5.3. *Suppose that $a_1,\ldots,a_\ell\in\mathcal{LE}$ satisfy $t^\varepsilon\prec a_i(t)\ll t$ for some $\varepsilon>0$, and the same is true for the functions $a_i(t)-a_j(t)$ for $i\neq j$.*

Then the factor $\mathcal{Z}$ is characteristic for the scheme $\{[a_1(n)],\ldots,[a_\ell(n)]\}$.

Proof. Let us first remark that in order to carry out our argument it will be convenient to work with a Hardy field H that contains $\mathcal{LE}$ and such that the set $H^+=\{a\in H\colon a(t)\to\infty\}$ is closed under composition and compositional inversion ($\mathcal{LE}$ does not have this last property). Such a Hardy field exists, in fact as mentioned in [11], a constructive example is the field of germs at ∞ of Pfaffian functions, which we denote by $\mathcal{P}$.[10] (We are not going to make use of the exact structure of $\mathcal{P}$.) Our goal is to show that if the functions $a_1,\ldots,a_\ell\in\mathcal{P}$ satisfy the stated assumptions, and one of the functions $f_1,\ldots,f_\ell$ is orthogonal to the nilfactor $\mathcal{Z}$, then

$$\lim_{N\to\infty}\Big\|\mathbb{E}_{1\le n\le N}T^{[a_1(n)]}f_1\cdot T^{[a_2(n)]}f_2\cdot\ldots\cdot T^{[a_\ell(n)]}f_\ell\Big\|_{L^2(\mu)}=0. \tag{51}$$

Without loss of generality we can assume that the function f_1 is orthogonal to $\mathcal{Z}$.

We are going to use induction on ℓ to show the following: If $(X,\mathcal{X},\mu,T)$ is a system, f_1 satisfies $\|f_1\|_\infty\le 1$, and the functions $a_1,\ldots,a_\ell\in\mathcal{P}$ satisfy the stated assumptions, then

$$\limsup_{N\to\infty}\sup_{\|f_0\|_\infty,\|f_2\|_\infty,\ldots,\|f_\ell\|_\infty\le 1}\mathbb{E}_{1\le n\le N}\Big|\int f_0\cdot T^{[a_1(n)]}f_1\cdot T^{[a_2(n)]}f_2\cdot\ldots\cdot T^{[a_\ell(n)]}f_\ell\, d\mu\Big|\ll_{\ell,a_1,\ldots,a_\ell}|\!|\!|f_1|\!|\!|_{2\ell}. \tag{52}$$

We also claim an analogous estimate with f_i in place of f_1 for $i=0,2,\ldots,\ell$. We leave it to the reader to verify that such estimates follow from (52) (for $i=2,\ldots,\ell$ by symmetry, for $i=0$ we factor out the transformation $T^{[a_1(n)]}$ and work with the resulting averages, the precise argument is very similar to the one given in the beginning of the proof of Lemma 4.7).

[10] This is defined inductively as follows: We let $\mathcal{P}_1$ be the set of all $f\in C^\infty(\mathbb{R}_+)$ that satisfy $f'=p(t,f)$ for some $p\in\mathbb{Z}[t_0,t_1]$, and for $k\ge 2$ we let $\mathcal{P}_k$ be the set of all $f\in C^\infty(\mathbb{R}_+)$ that satisfy $f'=p(t,f_1,\ldots,f_{k-1},f)$ for some $p\in\mathbb{Z}[t_0,t_1,\ldots,t_k]$ and $f_i\in\mathcal{P}_i$. Then $\mathcal{P}$ is the set of germs of functions in $\bigcup_{k\in\mathbb{N}}\mathcal{P}_k$.

First we verify that (52) implies (51). We can assume that $\|f_i\|_\infty \leq 1$ for $i=1,\ldots,\ell$. Since f_1 is orthogonal to $\mathcal{Z}$, we have $|||f_1|||_{2\ell}=0$, and as a consequence the limsup in (52) is 0. Using (52) with the conjugate of the function $\mathbb{E}_{1\leq n\leq N} T^{[a_1(n)]}f_1\cdot T^{[a_2(n)]}f_2\cdot\ldots\cdot T^{[a_\ell(n)]}f_\ell$ in place of the function f_0 (for every $N\in\mathbb{N}$), and removing the norms we get (51).

We proceed now to prove (52) by induction. Suppose first that $\ell=1$. If $a_1(t)\prec t$, then we deduce from Lemma 5.1 that

$$\lim_{N\to\infty}\Big(\sup_{\|f_0\|_\infty\leq 1}\mathbb{E}_{1\leq n\leq N}\Big|\int f_0\cdot T^{[a_1(n)]}f_1\ d\mu\Big| - \sup_{\|f_0\|_\infty\leq 1}\mathbb{E}_{1\leq n\leq N}\Big|\int f_0\cdot T^n f_1\ d\mu\Big|\Big)=0. \tag{53}$$

Equation (52) now follows by combining (53) and the estimate in Lemma 5.2. If $a_1(t)\sim t$, then $a_1(t)=\alpha t+e(t)$ for some nonzero $\alpha\in\mathbb{R}$ and $e(t)\prec t$. Assuming that $\alpha>0$ (the other case can be treated similarly), one sees that limit in (52) is bounded by a constant (one can use $[\alpha]+1$ if $\alpha>1$ and 2 if $\alpha\leq 1$) times the quantity

$$\limsup_{N\to\infty}\sup_{\|f_0\|_\infty\leq 1}\mathbb{E}_{1\leq n\leq N}\Big|\int f_0\cdot T^n f_1\ d\mu\Big|.$$

Combining this with Lemma 5.2 gives the advertised estimate.

Suppose now that $\ell\geq 2$ and the statement holds for $\ell-1$. We first claim that when proving (52) we can assume that the function $a_1(t)$ has maximal growth. Indeed, if $a_1(t)\prec a_i(t)$ for some $i\in 2,\ldots,\ell$, then we can factor out the transformation $T^{[a_i(n)]}$ and work with the resulting average. We omit the details since the argument is very similar to the one given in the beginning of the proof of Lemma 4.7.

We consider two cases:

Case 1. Suppose that $a_1(t)\sim t$. We can assume that for some $r\in\{1,\ldots,\ell\}$ we have $a_1(t)=\alpha_1 t+b_1(t),\ldots,a_r(t)=\alpha_r t+b_r(t)$ where α_i are non-zero real numbers, $b_i(t)\prec t$, and $a_{r+1}(t),\ldots,a_\ell(t)\prec t$.

We choose functions $f_{i,N}$ with $\|f_{i,N}\|_\infty\leq 1$ for $i=0,2,\ldots,\ell$, such that the value of the averages in (52) is $1/N$ close to the supremum. Using the Cauchy-Schwarz inequality we see that (52) follows if we show that

$$\limsup_{N\to\infty}\mathbb{E}_{1\leq n\leq N}\Big|\int f_{0,N}\cdot T^{[a_1(n)]}f_1\cdot T^{[a_2(n)]}f_{2,N}\ldots T^{[a_\ell(n)]}f_{\ell,N}\ d\mu\Big|^2\ll_{\ell,a_1,\ldots,a_\ell}|||f_1|||_{2\ell}^2. \tag{54}$$

Notice that the limit in (54) is equal to

$$A=\limsup_{N\to\infty}\mathbb{E}_{1\leq n\leq N}\int F_{0,N}\cdot S^{[a_1(n)]}F_1\cdot S^{[a_2(n)]}F_{2,N}\cdot\ldots\cdot S^{[a_\ell(n)]}F_{\ell,N}\ d(\mu\times\mu) \tag{55}$$

where

$$S=T\times T,\quad F_1=f_1\otimes\overline{f}_1,\quad\text{and}\quad F_{i,N}=f_{i,N}\otimes\overline{f}_{i,N}\ \text{ for }\ i=0,2,\ldots,\ell.$$

Using first the Cauchy Schwarz inequality, and then Lemma 4.6, we see that

$$|A|^2\leq 4\limsup_{H\to\infty}\mathbb{E}_{1\leq h\leq H}A_h \tag{56}$$

where

$$A_h=\limsup_{N\to\infty}\mathbb{E}_{1\leq n\leq N}\Big|\int S^{[a_1(n+h)]}F_1\cdot S^{[a_2(n+h)]}F_{2,N}\cdot\ldots\cdot S^{[a_\ell(n+h)]}F_{\ell,N}\cdot$$
$$S^{[a_1(n)]}\overline{F}_1\cdot S^{[a_2(n)]}F_{2,N}\cdot\ldots\cdot S^{[a_\ell(n)]}\overline{F}_{\ell,N}\ d(\mu\times\mu)\Big|.$$

We factor out the transformation $S^{[a_1(n)]}$. For $h\in\mathbb{N}$ fixed and large $n\in\mathbb{N}$, using that $b_i(n+h)-b_i(n)\to 0$ for $i=1,\ldots,r$ (since $b_i(t)\prec t$), and $a_i(n+h)-a_i(n)\to 0$ for $i=r+1,\ldots,\ell$ (since $a_i(t)\prec t$), we get the identities

$$\begin{aligned}
[a_i(n+h)]-[a_1(n)] &= [a_i(n)-a_1(n)]+[\alpha_i h]+e_i(h,n) \quad \text{for } i=1,\ldots,r,\\
[a_i(n+h)]-[a_1(n)] &= [a_i(n)-a_1(n)]+e_i(h,n) \quad \text{for } i=r+1,\ldots,\ell,\\
[a_i(n)]-[a_1(n)] &= [a_i(n)-a_1(n)]+\tilde{e}_i(h,n) \quad \text{for } i=2,\ldots,\ell,
\end{aligned}$$

where the error terms $e_i(h,n)$ and $\tilde{e}_i(h,n)$ take values in the set $\{0,\pm1,\pm2\}$.

We find that

$$A_h=\limsup_{N\to\infty}\mathbb{E}_{1\leq n\leq N}\Big|\int \tilde{F}_{1,h,n}\cdot S^{[a_2(n)-a_1(n)]}\tilde{F}_{2,h,n,N}\cdot\ldots\cdot S^{[a_\ell(n)-a_1(n)]}\tilde{F}_{\ell,h,n,N}\ d(\mu\times\mu)\Big|$$

where $\tilde{F}_{1,h,n}=T^{[\alpha_1 h]+e_1(h,n)}F_1\cdot\overline{F_1}$, $\tilde{F}_{i,h,n,N}=T^{[\alpha_i h]+e_i(h,n)}F_{i,N}\cdot T^{\tilde{e}_i(h,n)}\overline{F_{i,N}}$ for $i=2,\ldots,r$, and $\tilde{F}_{i,h,n,N}=T^{e_i(h,n)}F_{i,N}\cdot T^{\tilde{e}_i(h,n)}\overline{F_{i,N}}$ for $i=r+1,\ldots,\ell$. Next notice that for every fixed $h\in\mathbb{N}$ we can partition the integers into a finite number of sets, that depend only on ℓ, where all sequences $e_i(h,n),\tilde{e}_i(h,n)$ are constant. It follows that for every $h\in\mathbb{N}$ there exists $i\in\{0,\pm1,\pm2\}$ such that

$$A_h\ll_\ell \limsup_{N\to\infty}\sup_{\|F_2\|_\infty,\ldots,\|F_\ell\|_\infty\leq 1}\mathbb{E}_{1\leq n\leq N}\Big|\int \tilde{F}_{1,h,i}\cdot S^{[a_2(n)-a_1(n)]}F_2\cdot\ldots\cdot S^{[a_\ell(n)-a_1(n)]}F_\ell\ d(\mu\times\mu)\Big|$$

where $\tilde{F}_{1,h,i}=T^{[\alpha_1 h]+i}F_1\cdot\overline{F_1}$. Since the functions $a_2(t)-a_1(t),\ldots,a_\ell(t)-a_1(t)$ satisfy the assumptions of the induction hypothesis, it follows that

$$A_h\ll_{\ell,a_2-a_1,\ldots,a_\ell-a_1}\max_{i=0,\pm1,\pm2}|||\tilde{F}_{1,h,i}|||_{2(\ell-1)}. \tag{57}$$

($|||\tilde{F}_{1,h,i}|||_k$ is constructed using the system $(X\times X,\mathcal{X}\times\mathcal{X},S,\mu\times\mu)$.) Since

$$\tilde{F}_{1,h,i}=T^{[\alpha_1 h]+i}F_1\cdot\overline{F_1}=T^{[\alpha_1 h]+i}(f_1\otimes\overline{f_1})\cdot(\overline{f_1}\otimes f_1)=(T^{[\alpha_1 h]+i}f_1\cdot\overline{f_1})\otimes(\overline{T^{[\alpha_1 h]+i}f_1\cdot\overline{f_1}}),$$

and (21) gives

$$|||f\otimes\overline{f}|||_k=|||f|||_{k+1}^2$$

for every $f\in L^\infty(\mu)$ and $k\in\mathbb{N}$, we conclude that

$$|||\tilde{F}_{1,h,i}|||_{2(\ell-1)}=|||T^{[\alpha_1 h]+i}f_1\cdot\overline{f_1}|||_{2\ell-1}^2. \tag{58}$$

Combining (57) and (58) we get

$$\limsup_{H\to\infty}\mathbb{E}_{1\leq h\leq H}A_h\ll_{\ell,a_2-a_1,\ldots,a_\ell-a_1}\max_{i=1,..,5}\limsup_{H\to\infty}\mathbb{E}_{1\leq h\leq H}|||T^{[\alpha_1 h]+i}f_1\cdot\overline{f_1}|||_{2\ell-1}^2. \tag{59}$$

Finally, one sees that

$$\limsup_{H\to\infty}\mathbb{E}_{1\leq h\leq H}|||T^{[\alpha_1 h]+i}f_1\cdot\overline{f_1}|||_{2\ell-1}^2\leq([\alpha_1]+1)\limsup_{H\to\infty}\mathbb{E}_{1\leq h\leq H}|||T^h f_1\cdot\overline{f_1}|||_{2\ell-1}^2, \tag{60}$$

and using Hölder's inequality we get

$$\limsup_{H\to\infty}\mathbb{E}_{1\leq h\leq H}|||T^h f_1\cdot\overline{f_1}|||_{2\ell-1}^2\leq\limsup_{H\to\infty}\Big(\mathbb{E}_{1\leq h\leq H}|||T^h f_1\cdot\overline{f_1}|||_{2\ell-1}^{2^{2\ell-1}}\Big)^{\frac{1}{2^{2\ell-2}}}=|||f_1|||_{2\ell}^4. \tag{61}$$

(The last equality follows from (20).) Combining (56), (59), (60), and (61), we deduce that

$$|A|\ll_{\ell,a_1,\ldots,a_\ell}|||f_1|||_{2\ell}^2.$$

This proves (54) and completes the induction step in the case where $a_1(t)\sim t$.

Case 2. It remains to deal with the case $a_1(t) \prec t$. For $i = 1, \dots, \ell$, we write $a_i(t) = \tilde{a}_i(a_1(t))$ where $\tilde{a}_i(t) = a_i(a_1^{-1}(t))$. Notice that $\tilde{a}_i(t)$ is an element of $\mathcal{P}$ since $\mathcal{P}$ is closed under composition and compositional inversion. Keeping in mind that $a_1(t)$ has maximal growth and using Lemma 2.7 and Proposition 2.11 in [7], we get for some $\varepsilon > 0$ that $t^\varepsilon \prec \tilde{a}_i(t) \ll t$ for $i = 1, \dots, \ell$ and $t^\varepsilon \prec \tilde{a}_i(t) - \tilde{a}_j(t) \ll t$ for $i \neq j$. Hence, the functions $\tilde{a}_1(t) = t, \tilde{a}_2(t), \dots, \tilde{a}_\ell(t)$ satisfy the assumptions of the induction hypothesis.

Using Lemma 2.7 in [7] we get for $i = 2, \dots, \ell$ that

$$[a_i(n)] = [\tilde{a}_i([a_1(n)])] + e_i(n)$$

where the error terms $e_i(n)$ take values on a finite set. Therefore, the left hand side in (52) is equal to

$$\limsup_{N\to\infty} \sup_{\|f_0\|_\infty, \|f_2\|_\infty, \dots, \|f_\ell\|_\infty \leq 1} \mathbb{E}_{1\leq n\leq N} \Big| \int f_0 \cdot T^{[a_1(n)]} f_1 \cdot T^{[\tilde{a}_2([a_1(n)])]+e_2(n)} f_2 \cdot \ldots \cdot T^{[\tilde{a}_\ell([a_1(n)])]+e_\ell(n)} f_\ell \, d\mu \Big|.$$

Since the error terms $e_i(n)$ take values on a finite set, say with cardinality K (with K depending on the a_i's only), one sees that the previous limit is bounded by $K^{\ell-1}$ times

(62)
$$\limsup_{N\to\infty} \sup_{\|f_0\|_\infty, \|f_2\|_\infty, \dots, \|f_\ell\|_\infty \leq 1} \mathbb{E}_{1\leq n\leq N} \Big| \int f_0 \cdot T^{[a_1(n)]} f_1 \cdot T^{[\tilde{a}_2([a_1(n)])]} f_2 \cdot \ldots \cdot T^{[\tilde{a}_\ell([a_1(n)])]} f_\ell \, d\mu \Big|.$$

Since $t^\varepsilon \prec a_1(t) \prec t$, we can apply Lemma 5.1 for the sequence of real numbers $(V_N(n))_{N,n\in\mathbb{N}}$ defined by

$$V_N(n) = \Big| \int f_0 \cdot T^{[a_1(n)]} f_{1,N} \cdot T^{[\tilde{a}_2([a_1(n)])]} f_{2,N} \cdot \ldots \cdot T^{[\tilde{a}_\ell([a_1(n)])]} f_{\ell,N} \, d\mu \Big|,$$

where the functions $f_{i,N}$ are chosen so that the average in (62) gets $1/N$ close to the supremum. We get that the limit in (62) is equal to

$$\limsup_{N\to\infty} \sup_{\|f_0\|_\infty, \|f_2\|_\infty, \dots, \|f_\ell\|_\infty \leq 1} \mathbb{E}_{1\leq n\leq N} \Big| \int f_0 \cdot T^n f_1 \cdot T^{[\tilde{a}_2(n)]} f_2 \cdot \ldots \cdot T^{[\tilde{a}_\ell(n)]} f_\ell \, d\mu \Big|.$$

Since the functions $t, \tilde{a}_2(t), \dots, \tilde{a}_\ell(t)$ satisfy the assumptions of the induction hypothesis, we are reduced to Case 1. This completes the induction step and the proof. □

5.2. **The general case.** We are going to use an inductive method analogous to the one used in Section 4.3 for polynomial families. First, following [7], we introduce a notion of complexity that is suitable for our current setup.

We remind the reader that $\mathcal{G} = \{a \in C(\mathbb{R}_+) \colon t^{k+\varepsilon} \prec a(t) \prec t^{k+1}$ for some $k \geq 0$ and $\varepsilon > 0\}$. We say that a family $\mathcal{F} = \{a_1(t), \dots a_\ell(t)\}$ of functions in $\mathcal{LE}$ is *"nice"* if $a_i(t) \in \mathcal{G}$ and $a_i - a_j \in \mathcal{G}$ for $i \neq j$. Given any such "nice" family we associate a vector $(d, n_d, \dots, n_1, n_0)$, called the *type of* $\mathcal{F}$, with non-negative integer entries, as follows: For every non-negative integer i let

$$\mathcal{F}_i = \{a \in \mathcal{F} \colon t^i \prec a(t) \prec t^{i+1}\}.$$

We say that two functions $a, b \in \mathcal{F}_i$ are equivalent if $a(t) - b(t) \prec t^i$, and we define n_i to be the number of non-equivalent elements of $\mathcal{F}_i$. We denote the maximum integer i for which $n_i \neq 0$ by d. We order the set of all possible types lexicographically, meaning, $(d, w_d, \dots, w_1) > (d', w'_d, \dots, w'_1)$ if and only if in the first instance where the two vectors disagree the coordinate of the first vector is greater than the coordinate of the second vector.

Example 2. If $\mathcal{F} = \{t^{1/3}, t^{5/2}, t^{5/2} + t^{1/2}, t^{5/2} + t^{7/3}\}$, then the second and third functions are equivalent and all the other functions are non-equivalent. Hence, the type of $\mathcal{F}$ is $(2, 2, 0, 1)$.

Given a "nice" family of functions $\mathcal{F} = \{a_1(t), \dots a_\ell(t)\}$, a positive integer $h \in \mathbb{N}$, and $a \in \mathcal{F}$, we form a new family $\mathcal{F}(a(t), h)$ as follows: We start with the family of polynomials

$$\{a_1(t+h) - a(t), \dots, a_\ell(t+h) - a(t), a_1(t) - a(t), \dots, a_\ell(t) - a(t)\},$$

and successively remove the smallest number of functions so that the remaining set consists of unbounded functions whose pairwise differences are also unbounded. Then for every large h the family $\mathcal{F}(a, h)$ is also "nice" (if non-empty). The function $a_i(t+h) - a(t)$ will be removed if and only if $a_i(t) \prec t$, and the function $a_i(t) - a(t)$ will be removed if and only if $a(t) = a_i(t)$.

Example 3. If $\mathcal{F} = \{t^{1/3}, t^{1/2}, t^{3/2}\}$ and $a(t) = t^{1/3}$, then we start with the family of polynomials

$$\{(t+h)^{1/3} - t^{1/3}, (t+h)^{1/2} - t^{1/3}, (t+h)^{3/2} - t^{1/3}, t^{1/3} - t^{1/3}, t^{1/2} - t^{1/3}, t^{3/2} - t^{1/3}\}$$

and remove the first, second, and fourth functions to get

$$\mathcal{F}(t^{1/3}, h) = \{(t+h)^{3/2} - t^{1/3}, t^{1/2} - t^{1/3}, t^{3/2} - t^{1/3}\}.$$

Notice that the family $\mathcal{F}$ has type $(1, 1, 2)$, and the family $\mathcal{F}(t^{1/3}, h)$ has smaller type, namely, $(1, 1, 1)$.

To prove Theorem 2.9 we are going to use induction on the type of the family of functions involved. In order to carry out the inductive step we will use the following:

Lemma 5.4. *Let $\mathcal{F} = \{a_1(t), \dots a_\ell(t)\}$ be a "nice" family of functions. Suppose that $a_1(t) \succ t$, and $a_1(t)$ has maximal growth rate within $\mathcal{F}$, meaning $a_i(t) \ll a_1(t)$ for $i = 2, \dots, \ell$.*

Then it is possible to choose $a \in \mathcal{F}$ such that for every large h the family $\mathcal{F}(a(t), h)$ is "nice", has type smaller than that of $\mathcal{F}$, and the function $a_1(t+h) - a(t)$ has maximal growth rate within the family $\mathcal{F}(a(t), h)$.

Remark. Since $a_1(t) \succ t$, no-matter what the choice of the function $a(t)$ will be, the function $a_1(t+h) - a(t)$ is going to be an element of the family $\mathcal{F}(a(t), h)$ for every large h.

Proof. Suppose first that $a_i(t) \prec a_1(t)$ for some $i \in \{2, \dots, \ell\}$. Let i_0 be such that the function $a_{i_0}(t)$ has the minimal growth (meaning $a_{i_0}(t) \ll a_i(t)$ for $i = 1, \dots, \ell$). Then $a(t) = a_{i_0}(t)$ has the advertised property.

Otherwise, $a_i(t) \sim a_1(t)$ for $i = 1, \dots, \ell$, in which case, for $i = 2, \dots, \ell$, we have $a_i(t) = \alpha_i a_1(t) + b_i(t)$, for some non-zero real numbers $\alpha_2, \dots, \alpha_\ell$, and functions $b_i(t)$ with $b_i(t) \prec a_1(t)$. If $\alpha_{i_0} \neq 1$ for some $i_0 \in \{2, \dots, \ell\}$, then $a(t) = a_{i_0}(t)$ has the advertised property. If $\alpha_i = 1$ for $i = 2, \dots, \ell$, let i_0 be such that the function $b_{i_0}(t)$ has maximal growth. Then $a(t) = a_{i_0}(t)$ has the advertised property. This completes the proof. $\square$

We are now ready to give the proof of Theorem 2.9. We recall its statement for convenience.

Theorem 5.5. *Suppose that $\{a_1(t), \dots, a_\ell(t)\}$ is a "nice" family of functions in $\mathcal{LE}$.*

Then the factor $\mathcal{Z}$ is characteristic for the family $\{[a_1(n)], \dots, [a_\ell(n)]\}$.

Proof. Arguing as in the beginning of the proof of Proposition 4.7 we see that it suffices to show the following: If $(X, \mathcal{X}, \mu, T)$ is a measure preserving system, $f_1 \in L^\infty(\mu)$ is orthogonal

to the nilfactor $\mathcal{Z}$, the family of functions $\mathcal{F} = \{a_1(t), \ldots, a_\ell(t)\}$ is nice, and the function $a_1(t)$ has maximal growth within $\mathcal{F}$, then

$$(63) \quad \lim_{N\to\infty} \sup_{\|f_0\|_\infty, \|f_2\|_\infty, \ldots, \|f_\ell\|_\infty \le 1} \mathbb{E}_{1\le n\le N} \Big| \int f_0 \cdot T^{[a_1(n)]} f_1 \cdot T^{[a_2(n)]} f_2 \cdot \ldots \cdot T^{[a_\ell(n)]} f_\ell \, d\mu \Big| = 0.$$

We are going to use induction on the type of the family $\{a_1(t), \ldots, a_\ell(t)\}$ to verify this statement.

Proposition 5.3 shows that the result holds for $d = 0$. Suppose now that $d \ge 1$, and our statement holds for all families with type smaller than $(d, n_d, \ldots, n_0)$. Let $\mathcal{F} = \{a_1(t), \ldots, a_\ell(t)\}$ be a family with type $(d, n_d, \ldots, n_0)$ (then $a_1(t) \succ t$ since $a_1(t)$ has maximal growth rate in $\mathcal{F}$).

Choosing functions $f_{i,N}$ with $\|f_{i,N}\|_\infty \le 1$, so that the average in (63) is $1/N$ close to the supremum, and using the Cauchy-Schwarz inequality, we see that it suffices to show that

$$\lim_{N\to\infty} \mathbb{E}_{1\le n\le N} \Big| \int f_{0,N} \cdot T^{[a_1(n)]} f_1 \cdot T^{[a_2(n)]} f_{2,N} \ldots T^{[a_\ell(n)]} f_{\ell,N} \, d\mu \Big|^2 = 0.$$

Equivalently, it suffices to show that

$$\lim_{N\to\infty} \mathbb{E}_{1\le n\le N} \int F_{0,N} \cdot S^{[a_1(n)]} F_1 \cdot S^{[a_2(n)]} F_{2,N} \cdot \ldots \cdot S^{[a_\ell(n)]} F_{\ell,N} \, d(\mu \times \mu) = 0$$

where $S = T \times T$, $F_1 = f_1 \otimes \overline{f}_1$, and $F_{i,N} = f_{i,N} \otimes \overline{f}_{i,N}$, for $i = 0, 2, \ldots, \ell$. Using the Cauchy-Schwarz inequality, and then Lemma 4.6, we reduce matters to showing for every large h that

$$\lim_{N\to\infty} \Big| \mathbb{E}_{1\le n\le N} \int S^{[a_1(n+h)]} F_1 \cdot S^{[a_2(n+h)]} F_{2,N} \cdot \ldots \cdot S^{[a_\ell(n+h)]} F_{\ell,N} \cdot$$
$$S^{[a_1(n)]} \overline{F}_1 \cdot S^{[a_2(n)]} \overline{F}_{2,N} \cdot \ldots \cdot S^{[a_\ell(n)]} \overline{F}_{\ell,N} \, d(\mu \times \mu) \Big| = 0.$$

Factoring out the transformation $S^{[a(n)]}$ where $a(t) \in \{a_1(t), \ldots, a_\ell(t)\}$ is as in Lemma 5.4, we see that it suffices to show that for every large h we have

$$\lim_{N\to\infty} \mathbb{E}_{1\le n\le N} \Big| \int S^{[a_1(n+h)-a(n)]+e_1(n)} F_1 \cdot S^{[a_2(n+h)-a(n)]+e_2(n)} F_{2,N} \cdot \ldots \cdot S^{[a_\ell(n+h)-a(n)]+e_\ell(n)} F_{\ell,N} \cdot$$
$$S^{[a_1(n)-a(n)]+e_{\ell+1}(n)} \overline{F}_1 \cdot S^{[a_2(n)-a(n)]+e_{\ell+2}(n)} \overline{F}_{2,N} \cdot \ldots \cdot S^{[a_\ell(n)-a(n)]+e_{2\ell}(n)} \overline{F}_{\ell,N} \, d(\mu \times \mu) \Big| = 0$$

for some error terms $e_i(n)$ that take values in $\{0, 1\}$. Therefore, it suffices to show that for every large h we have

(64)

$$\lim_{N\to\infty} \sup_{\|F_2\|_\infty, \ldots, \|F_{2\ell}\|_\infty \le 1} \mathbb{E}_{1\le n\le N} \Big| \int S^{[a_1(n+h)-a(n)]} F_1 \cdot S^{[a_2(n+h)-a(n)]} F_2 \cdot \ldots \cdot S^{[a_\ell(n+h)-a(n)]} F_\ell \cdot$$
$$S^{[a_1(n)-a(n)]} F_{l+1} \cdot S^{[a_2(n)-a(n)]} F_{l+2} \cdot \ldots \cdot S^{[a_\ell(n)-a(n)]} F_{2\ell} \, d(\mu \times \mu) \Big| = 0.$$

Next notice that if $a_i(t) \prec t$ for some $i \in \{2, \ldots, \ell\}$, then $a_i(n+h) - a_i(n) \to 0$. Therefore, for our purposes we can practically assume that $[a_i(n+h) - a(n)] = [a_i(n) - a(n)]$ for all n, and for those values of i we can write

$$S^{[a_i(n+h)-a(n)]} F_i \cdot S^{[a_i(n)-a(n)]} \overline{F}_i = S^{[a_i(n)-a(n)]} |F_i|^2.$$

After doing this reduction, for every large h the iterates that appear in the integral in (64) involve functions that belong to the family $\mathcal{F}(a(t), h)$ (defined at the beginning of the current

subsection). This family is "nice" and by Lemma 5.4 has type smaller than that of $\mathcal{F}$. Furthermore, one of these iterates appearing in this reduced form is $S^{[a_1(n+h)-a(n)]}F_1$ and the function $a_1(n+h)-a(n)$ has maximal growth in $\mathcal{F}(a(t),h)$. Since f_1 is orthogonal to the factor $\mathcal{Z}(T)$, we have that $F_1 = f_1 \otimes \overline{f}_1$ is orthogonal to the factor $\mathcal{Z}(S)$. Therefore, the induction hypothesis applies, and gives that the limit in (64) is 0 for every large h. This completes the induction and the proof. $\square$

6. Proof of the convergence and the recurrence results

In this section we combine Theorems 2.4 and 2.9 and the equidistribution results from [14] to prove the convergence and recurrence results stated in Sections 2.1 and 2.2.

6.1. **Convergence results.** We prove the convergence results stated in Sections 2.1 and 2.2. To prove Theorem 2.1 we will need the following result:

Theorem 6.1 (**F.** [14]). *Suppose that the function $a \in \mathcal{H}$ has polynomial growth and satisfies one of the three conditions stated in Theorem 2.1.*

Then for every nilmanifold $X = G/\Gamma$, $F \in C(X)$, $b \in G$, and $x \in X$, the following limit exists $\lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} F(b^{[a(n)]}x)$.

Proof of Theorem 2.1. The necessity of the conditions was proved in [12] by using examples of rational rotations on the circle.

To show that the three stated conditions are sufficient for convergence we start by using Theorem 2.4. We get that the nilfactor $\mathcal{Z}$ is characteristic for the corresponding multiple ergodic averages. Using an ergodic decomposition argument and Theorem 3.3, we deduce that it suffices to prove convergence when our system is an inverse limit of nilsystems. A standard approximation argument allows us to finally reduce matters to nilsystems.

Let $(X = G/\Gamma, \mathcal{G}/\Gamma, m_X, T_b)$ be a nilsystem and $F_1, \dots, F_\ell \in L^\infty(m_X)$. Our goal is to show that if the function $a \in \mathcal{H}$ satisfies one of the three stated conditions, then the limit

$$\lim_{N\to\infty} \mathbb{E}_{1\leq n\leq N} F_1(b^{[a(n)]}x)\cdot F_2(b^{2[a(n)]}x)\cdot \ldots \cdot F_\ell(b^{\ell[a(n)]}x) \tag{65}$$

exists in $L^2(m_X)$. By density, we can assume that the functions $F_1, \dots, F_\ell$ are continuous. In this case we claim that the limit in (65) exists for every $x \in X$. Indeed, applying Theorem 6.1 to the nilmanifold X^k, the nilrotation $\tilde{b} = (b, b^2, \dots, b^\ell)$, the point $\tilde{x} = (x, x, \dots, x)$, and the function $\tilde{F}(x_1, \dots, x_\ell) = F_1(x_1)\cdot F_2(x_2)\cdots F_\ell(x_l)$, we get that the limit

$$\lim_{N\to\infty} \mathbb{E}_{1\leq n\leq N} \tilde{F}(\tilde{b}^{[a(n)]}\tilde{x})$$

exists. This implies that the limit in (65) exists for every $x \in X$ and completes the proof. $\square$

To prove Theorem 2.2 we will need the following result:

Theorem 6.2 (**F.** [14]). *Let $a \in \mathcal{H}$ have at most polynomial growth and satisfy $|a(t) - cp(t)| \succ \log t$ for every $c \in \mathbb{R}$ and $p \in \mathbb{Z}[t]$.*

Then for every nilmanifold $X = G/\Gamma$, $b \in G$, and $x \in X$, the sequence $(b^{[a(n)]}x)_{n\in\mathbb{N}}$ is equidistributed in the nilmanifold $\overline{(b^n x)}_{n\in\mathbb{N}}$.

Proof of Theorem 2.2. Arguing as in the proof of Theorem 2.1 we reduce matters to showing the following: Let $a \in \mathcal{H}$ satisfy the assumptions of our theorem, $X = G/\Gamma$ be a nilsystem,

$b \in G$ be a nilrotation, and $F_1, \dots, F_\ell \in C(X)$. Then for every $x \in X$ the limit in (65) exists and is equal to the limit

$$\lim_{N\to\infty} \mathbb{E}_{1\leq n\leq N} F_1(b^n x) \cdot F_2(b^{2n} x) \cdot \ldots \cdot F_\ell(b^{\ell n} x).$$

Keeping the same notation as in the proof of Theorem 2.2, we rewrite the desired identity as

$$\lim_{N\to\infty} \mathbb{E}_{1\leq n\leq N} \tilde{F}(\tilde{b}^{[a(n)]}\tilde{x}) = \lim_{N\to\infty} \mathbb{E}_{1\leq n\leq N} \tilde{F}(\tilde{b}^{n}\tilde{x}).$$

Using Theorem 6.2, and the fact that the sequence $(\tilde{b}^n\tilde{x})_{n\in\mathbb{N}}$ is equidistributed in the set $\overline{\{\tilde{b}^n\tilde{x}, n \in \mathbb{N}\}}$, we deduce that the last identity holds for every $\tilde{x} \in X^\ell$, completing the proof. $\square$

Next we prove Theorem 2.6 using the following result that will be established in Section 6.3.

Proposition 6.3. *Suppose that the functions $a_1, \dots, a_\ell \in \mathcal{LE} \cap \mathcal{G}$ have different growth rates. Then for every nilmanifold $X = G/\Gamma$, $b \in G$ ergodic, and $F_1, \dots, F_\ell \in C(X)$ we have*

$$\lim_{N\to\infty} \mathbb{E}_{1\leq n\leq N} F_1(b^{[a_1(n)]}x) \cdot \ldots \cdot F_\ell(b^{[a_\ell(n)]}x) = \int F_1 \, dm_X \cdot \ldots \cdot \int F_\ell \, dm_X$$

where the convergence takes place in $L^2(m_X)$.

Proof of Theorem 2.6. Using Theorem 2.9 and arguing as in the proof of Theorem 2.1 the result follows from Proposition 6.3. $\square$

Lastly, we prove Theorem 2.7. Its proof has a rather soft touch of ergodic theory; in fact it is an easy consequence of the following general statement:

Proposition 6.4. *Let $a_1, \dots, a_\ell \in \mathcal{LE} \cap \mathcal{G}$ have different growth rates and satisfy $a_i(t) \prec t$ for $i = 1, \dots, \ell$. Let $(X, \mathcal{X}, \mu)$ be a probability space, and for $i = 1, \dots, \ell$ let $(F_i(n))_{n\in\mathbb{N}}$ be sequences of functions in $L^\infty(\mu)$ with uniformly bounded norm such that the limits $\tilde{F}_i = \lim_{N-M\to\infty} \mathbb{E}_{M\leq n\leq N} F_i(n)$ exist in $L^2(\mu)$.*

Then

$$\lim_{N\to\infty} \mathbb{E}_{1\leq n\leq N} F_1([a_1(n)]) \cdot \ldots \cdot F_\ell([a_\ell(n)]) = \tilde{F}_1 \cdot \ldots \cdot \tilde{F}_\ell$$

where the limit is taken in $L^2(\mu)$.

Proof. As in the proof of Proposition 5.3 it will be convenient to work with the larger Hardy field $\mathcal{P}$ of Pfaffian functions which is closed under composition and compositional inversion.

We use induction on ℓ. For $\ell = 1$ the result follows from Lemma 5.1.

Suppose that the result holds for $\ell - 1$. Without loss of generality we can assume that $a_i(t) \prec a_\ell(t)$ for $i = 1, \dots, \ell - 1$. Assuming that $\tilde{F}_\ell = 0$ it suffices to prove that

$$\lim_{N\to\infty} \mathbb{E}_{1\leq n\leq N} F_1([a_1(n)]) \cdot \ldots \cdot F_\ell([a_\ell(n)]) = 0 \tag{66}$$

where the convergence takes place in $L^2(\mu)$. For $i = 1, \dots \ell - 1$ we let $\tilde{a}_i(t) = a_i(a_\ell^{-1}(t))$ which is again an element of $\mathcal{P}$. Since the functions $a_1(t), \dots, a_\ell(t)$ have different growth rates, belong to $\mathcal{G}$, and $a_\ell(t)$ has maximal growth, the same is the case for the functions $\tilde{a}_1(t), \dots, \tilde{a}_{\ell-1}(t), t$, and $\tilde{a}_i(t) \prec t$ for $i = 1, \dots, \ell - 1$. Keeping this in mind and using Lemma 2.7 in [7] we get that $t^\varepsilon \prec \tilde{a}_i(t) \prec t$ for some $\varepsilon > 0$. Furthermore, by Lemma 2.12 in [7] we have $[a_i(n)] = \tilde{a}_i([a_\ell(n)])$ for a set of $n \in \mathbb{N}$ of density 1. It follows from Lemma 5.1 that in order to verify (66) it suffices to show that

$$\lim_{N\to\infty} \mathbb{E}_{1\leq n\leq N} F_1([\tilde{a}_1(n)]) \cdot \ldots \cdot F_{\ell-1}([\tilde{a}_{\ell-1}(n)]) \cdot F_\ell(n) = 0. \tag{67}$$

Since for $i = 1, \dots, \ell - 1$ the functions $\tilde{a}_i(t)$ belong to some Hardy field and satisfy $\tilde{a}_i(t) \prec t$ we have that $\tilde{a}_i(t+1) - \tilde{a}_i(t) \to 0$. Using this, we see that there exists a sequence $(I_m)_{m\in\mathbb{N}}$ of non-overlapping intervals, with $|I_m| \to \infty$, and such that the sequences $[\tilde{a}_1(n)], \dots, [\tilde{a}_{\ell-1}(n)]$ are constant on every interval I_m (for technical reasons we can also assume that $|I_m| \prec m$). Then (67) follows if we show that

$$\lim_{M\to\infty} \mathbb{E}_{1\leq m\leq M}\big(G_m \cdot \mathbb{E}_{n\in I_m} F_\ell(n)\big) = 0, \tag{68}$$

where

$$G_m = F_1([\tilde{a}_1(n_m)]) \cdot \ldots \cdot F_{\ell-1}([\tilde{a}_{\ell-1}(n_m)]),$$

and n_m is any element of the interval I_m. Furthermore, since the functions G_m have uniformly bounded L^∞ norms, it suffices to show that

$$\lim_{M\to\infty} \mathbb{E}_{1\leq m\leq M} \left\|\mathbb{E}_{n\in I_m} F_\ell(n)\right\|_{L^2(\mu)} = 0. \tag{69}$$

Our assumption gives that

$$\lim_{m\to\infty} \left\|\mathbb{E}_{n\in I_m} F_\ell(n)\right\|_{L^2(\mu)} = 0,$$

which immediately implies (69). This completes the proof. $\square$

We deduce Theorem 2.7 from Proposition 6.4.

Proof of Theorem 2.7. By the (uniform) mean ergodic theorem we have

$$\lim_{N-M\to\infty} \left\|\mathbb{E}_{M\leq n\leq N} T_i^n f_i\right\|_{L^2(\mu)} = E(f_i|\mathcal{I}(T_i))$$

where the convergence takes place in $L^2(\mu)$. We can therefore apply Proposition 6.4 for the sequences $(F_i(n))_{n\in\mathbb{N}}$, $i = 1, \dots, \ell$, defined by $F_i(n) = T_i^n f_i$ to conclude the proof. $\square$

6.2. **Recurrence results.** We prove Theorem 2.3 using the following multiple recurrence result that will be handled later.

Proposition 6.5. *Let $a \in \mathcal{H}$ be of the form $a(t) = p(t)\alpha + e(t)$ for some $p \in \mathbb{Z}[t]$, $\alpha \in \mathbb{R}$, and $1 \prec e(t) \prec t$. Let $(X, \mathcal{X}, \mu, T)$ be a system and $f \in L^\infty(\mu)$ be non-negative and not almost everywhere zero.*

Then for every $\ell \in \mathbb{N}$ we have

$$\limsup_{N-M\to\infty} \mathbb{E}_{M\leq n\leq N} \int f \cdot T^{[a(n)]} f \cdot \ldots \cdot T^{\ell[a(n)]} f \; d\mu > 0. \tag{70}$$

Proof of Theorem 2.3. If $|a(t) - cp(t)| \succ \log t$ for every $c \in \mathbb{R}$ and $p \in \mathbb{Z}[t]$, then the result follows immediately by combining Theorem 2.2 and Furstenberg's multiple recurrence theorem ([18]). Furthermore, the case where $a(t) = cp(t) + e(t)$ for some $c \in \mathbb{R}$, $p \in \mathbb{Z}[t]$, and $e(t) \ll \log t$, is taken care by Proposition 6.5. This completes the proof. $\square$

Lastly, we deduce Theorem 2.8 from Theorem 2.6.

Proof of Theorem 2.8. Let $\mu = \int \mu_t d\lambda(t)$ be the ergodic decomposition of the measure μ. Using Theorem 2.6 for the ergodic systems $(X, \mathcal{X}, \mu_t, T)$ we get that

$$\lim_{N\to\infty} \mathbb{E}_{1\leq n\leq N} \mu(A \cap T^{-[a_1(n)]}A \cap T^{-[a_2(n)]}A \cap \cdots \cap T^{-[a_\ell(n)]}A) = \int (\mu_t(A))^{\ell+1} \; d\lambda(t).$$

Using Holder's inequality we see that the last integral is at least

$$\Big(\int \mu_t(A) \; d\lambda(t)\Big)^{\ell+1} = (\mu(A))^{\ell+1}.$$

This proves the advertised estimate and completes the proof. $\square$

6.3. **Proof of Proposition 6.3.** In the case where all the functions $a_1(t), \dots, a_\ell(t)$ have super-linear growth Proposition 6.3 is a direct consequence of the corresponding pointwise result in [14] (Theorem 1.3). The general case can be covered using a modification of an argument used in [14]. To avoid unnecessary repetition our proof will be rather sketchy.

We are going to use the following:

Proposition 6.6. *Suppose that the functions $a_1, \dots, a_\ell \in \mathcal{LE} \cap \mathcal{G}$ have different growth rates. Then for every nilmanifold $X = G/\Gamma$, $b \in G$, $x \in X$, and $F \in C(X^\ell)$, we have*

$$\lim_{R\to\infty} \lim_{N\to\infty} \mathbb{E}_{1\leq n\leq N} \Big| \mathbb{E}_{1\leq r\leq R} F(b^{[a_1(Rn+r)]}x, \dots, b^{[a_\ell(Rn+r)]}x) - \int F \ dm_{X_b^\ell} \Big| = 0$$

where $X_b = \overline{\{b^n x \colon n \in \mathbb{N}\}}$.

Sketch of Proof. Using a straightforward modification of the reduction argument of Section 5.2 in [14], we can reduce matters to showing that for every nilmanifold $X = G/\Gamma$, with G connected and simply connected, $b \in G$ ergodic, and $F \in C(X^\ell)$, we have

$$\lim_{R\to\infty} \lim_{N\to\infty} \mathbb{E}_{1\leq n\leq N} \Big| \mathbb{E}_{1\leq r\leq R} F(b^{[a_1(Rn+r)]}x, \dots, b^{[a_\ell(Rn+r)]}x) - \int F \ dm_{X^\ell} \Big| = 0,$$

This was verified while proving Proposition 5.3 in [14], completing the proof. $\square$

Proof of Proposition 6.3. For convenience of exposition we are going to assume that all the limits mentioned below exist; their existence will follow once our argument is completed.

Our goal is to show that for every nilmanifold $X = G/\Gamma$, $b \in G$ ergodic, and $F_1, \dots, F_\ell \in C(X)$ we have

$$\lim_{N\to\infty} \mathbb{E}_{1\leq n\leq N} F_1(b^{[a_1(n)]}x) \cdot \ldots \cdot F_\ell(b^{[a_\ell(n)]}x) = \int F_1 \ dm_X \cdot \ldots \cdot \int F_\ell \ dm_X$$

where the convergence takes place in $L^2(m_X)$.

If all the functions $a_1(t), \dots, a_\ell(t)$ are sub-linear, then the result follows from Theorem 7.3 in [7], and if all the functions are super-linear, then the result follows from Theorem 1.3 in [14]. If none of these is the case, we can assume that $a_1(t), \dots, a_m(t) \succ t$ and $a_{m+1}(t), \dots, a_\ell(t) \prec t$ for some $m \in \{1, \dots, \ell - 1\}$.

Let $F \in C(X^m)$ and $G \in C(X^{\ell-m})$. It suffices to show that

(71)
$$\lim_{N\to\infty} \mathbb{E}_{1\leq n\leq N} \big(F(b^{[a_1(n)]}x, \dots, b^{[a_m(n)]}x) \cdot G(b^{[a_{m+1}(n)]}x, \dots, b^{[a_\ell(n)]}x)\big) = \int F dm_{X^m} \int G dm_{X^{\ell-m}}$$

where the convergence takes place in $L^2(m_X)$. For every $R \in \mathbb{N}$ the limit in (71) is equal to

(72)
$$\lim_{N\to\infty} \mathbb{E}_{1\leq n\leq N} \Big(\mathbb{E}_{1\leq r\leq R} \big(F(b^{[a_1(nR+r)]}x, \dots, b^{[a_m(nR+r)]}x) \cdot G(b^{[a_{m+1}(nR+r)]}x, \dots, b^{[a_\ell(nR+r)]}x)\big)\Big).$$

Since the functions $a_{m+1}, \dots, a_\ell \in \mathcal{LE}$ are all sub-linear, we can show (see Lemma 2.12 in [7]) the following: for every $R \in \mathbb{N}$, for a set of $n \in \mathbb{N}$ of density 1, we have $[a_i(nR + r)] = [a_i(nR)]$ for $r = 1, \dots, R$ and $i = m + 1, \dots, \ell$. We deduce that the limit in (72) is equal to

(73) $$\lim_{N\to\infty} \mathbb{E}_{1\leq n\leq N} \Big(G(b^{[a_{m+1}(nR)]}x, \dots, b^{[a_\ell(nR)]}x) \cdot \mathbb{E}_{1\leq r\leq R} F(b^{[a_1(nR+r)]}x, \dots, b^{[a_m(nR+r)]}x)\Big).$$

Using Proposition 6.6 we see that the limit of the expression (73) as $R\to\infty$ is equal to

$$\lim_{R\to\infty}\lim_{N\to\infty}\mathbb{E}_{1\leq n\leq N}G(b^{[a_{m+1}(nR)]}x,\ldots,b^{[a_\ell(nR)]}x)\cdot\int F\;dm_{X^m}.$$

Since $b\in G$ is ergodic and the functions $a_{m+1},\ldots,a_\ell\in\mathcal{LE}\cap\mathcal{G}$ are all sub-linear and have different growth, by Proposition 6.4 in [7] we have (for every $R\in\mathbb{N}$) that

$$\lim_{N\to\infty}\mathbb{E}_{1\leq n\leq N}G(b^{[a_{m+1}(nR)]}x,\ldots,b^{[a_\ell(nR)]}x)=\int G\;dm_{X^{\ell-m}}$$

where the convergence takes place in $L^2(\mu)$. Combining all the previous facts we get (71). This completes the proof. $\square$

6.4. **Proof of Proposition 6.5.** First, we informally discuss the proof strategy of Proposition 6.5. When the function $a(t)$ is logarithmically close to a constant multiple of an integer polynomial the relevant multiple ergodic averages cannot be directly compared with Furstenberg's averages (because of the luck of equidistribution). To bypass this difficulty we work with an appropriate subsequence of the sequence $[a(n)]$. This subsequence is chosen so that it becomes possible to compare the corresponding multiple ergodic averages with those along the sequence $[n\alpha]$. For the latter averages positiveness follows easily from Furstenberg's multiple recurrence theorem, thus achieving our goal.

To carry out this plan we need a few preliminary results that enable us to carry out the aforementioned "comparison step".

We start with an equidistribution result on nilmanifolds. To prove the claimed uniformity we make use some quantitative equidistribution results (this is the only place in the present article where we make explicit use of such results).

Lemma 6.7. *Let $X=G/\Gamma$ be a connected nilmanifold, $b\in G$ be an ergodic nilrotation, and $p\in\mathbb{Z}[t]$ be non-constant.*

Then for every $F\in C(X)$ we have

$$\lim_{N-M\to\infty}\max_{x\in X}\left|\mathbb{E}_{M\leq n\leq N}F(b^{p(n)}x)-\int F\;dm_X\right|=0. \tag{74}$$

Proof. We argue by contradiction. Suppose that (74) fails for some connected nilmanifold X, ergodic $b\in G$, $p\in\mathbb{Z}[t]$ with $k=\deg(p)\geq 1$, and $F\in C(X)$. Then there exist $\delta>0$, $x_m\in X$, and sequences of positive integers $(n_m)_{m\in\mathbb{N}}$, $(N_m)_{m\in\mathbb{N}}$, with $N_m\to\infty$, such that the sequence $(b^{p(n_m+n)}x_m)_{1\leq n\leq N_m}$ is not δ-equidistributed in X for every $m\in\mathbb{N}$.

Then by Theorem 3.2 (suppose that δ is small enough so that the theorem applies) there exists a constant $M=M(X,\delta,k)$, and a sequence of quasi-characters ψ_m, with $\|\psi_m\|\leq M$, and such that

$$\left\|\psi_m(b^{p(n_m+n)}x_m)\right\|_{C^\infty[N_m]}\leq M \tag{75}$$

for every $m\in\mathbb{N}$. As explained in Section 3.2, the affine torus A of X can be identified with a finite dimensional torus $\mathbb{T}^l$. After making this identification, we have $\psi_m(t)=\kappa_m\cdot t$ for some non-zero $\kappa_m\in\mathbb{Z}^l$, and the nilrotation b induces a d-step unipotent affine transformation $T_b\colon\mathbb{T}^l\to\mathbb{T}^l$. Let $\pi(b)=(\beta_1\mathbb{Z},\ldots,\beta_s\mathbb{Z})$, where $\beta_i\in\mathbb{R}$, be the projection of b to the Kronecker factor of T_b (notice that s is bounded by the dimension of X). Since b acts ergodically on X the set $B=\{1,\beta_1,\ldots,\beta_s\}$ is rationally independent. For every $x\in\mathbb{T}^l$ the coordinates of $T_b^n x$

are polynomials of n, and so $\kappa_m \cdot T_b^n x$ is a polynomial of n. Moreover, the leading term of the polynomial $\kappa_m \cdot T_b^n x$ has the form $\gamma_m n^{\tilde{d}}$, where $1 \leq \tilde{d} \leq d$, and

$$\gamma_m = \frac{1}{k!} \sum_{i=1}^{s} r_{m,i}\beta_i, \quad \text{where } r_i \in \mathbb{Z} \text{ are not all zero and } |r_{m,i}| \leq c_1 \cdot M \tag{76}$$

for some constant c_1 that depends only on b.[11] Using this and the definition of $\|\cdot\|_{C^\infty[N]}$ (see (19)), it follows that

$$\left\| \psi(T_b^{p(n_m+n)} x_m) \right\|_{C^\infty[N_m]} \geq N_m^{k\tilde{d}} \|\gamma_m\|.$$

Combining this with (75) we get that

$$\|\gamma_m\| \leq \frac{M}{N_m^{k\tilde{d}}}. \tag{77}$$

Since by (76) we have a finite number of options for the irrational numbers γ_m, and $N_m \to \infty$, letting $m \to \infty$ we get a contradiction. This completes the proof. $\square$

Next we use the Lemma 6.7 to establish a key identity. We remark that if G is a connected and simply connected nilpotent Lie group, then there exists a unique continuous homomorphism $b \colon \mathbb{R} \to G$ with $b(1) = b$. For $b \in G$ and $t \in \mathbb{R}$, by b^t we mean the element $b(t) \in G$.

Lemma 6.8. *Let $X = G/\Gamma$ be a nilmanifold with G connected and simply connected, $\alpha \in \mathbb{R}$ be non-zero, and $b \in G$ be such that the sub-nilmanifold $\overline{\{(n\alpha\mathbb{Z}, b^{n\alpha}\Gamma), n \in \mathbb{N}\}}$ of $\mathbb{T} \times X$ is connected.*

Then for every $F \in C(X)$, Følner sequence $(\Phi_m)_{m\in\mathbb{N}}$ in $\mathbb{Z}$, and non-constant $p \in \mathbb{Z}[t]$ we have

$$\lim_{m\to\infty} \mathbb{E}_{n\in\Phi_m} F\big(b^{[p(n)\alpha+m\alpha]}\Gamma\big) = \lim_{N\to\infty} \mathbb{E}_{1\leq n\leq N} F(b^{[n\alpha]}\Gamma). \tag{78}$$

Proof. We first do some maneuvers that enable us to remove the integer parts and bring us to a point where Lemma 6.7 is applicable. We let $\tilde{X}$ be the nilmanifold $\mathbb{T} \times X$ which we identify with $(\mathbb{R} \times G)/(\mathbb{Z} \times \Gamma)$, set $\tilde{b} = (\alpha, b^\alpha)$, and define the function $\tilde{F}$ on $\tilde{X}$ by

$$\tilde{F}(t\mathbb{Z}, g\Gamma) = F(b^{-\{t\}} g\Gamma).$$

Notice that for every $t \in \mathbb{R}$ we have

$$\tilde{F}(\tilde{b}^t\tilde{\Gamma}) = F(b^{-\{\alpha t\}} b^{\alpha t}\Gamma) = F(b^{[\alpha t]}\Gamma) \tag{79}$$

where $\tilde{\Gamma} = \mathbb{Z} \times \Gamma$. (We caution the reader that the function $\tilde{F}$ is not continuous on $\tilde{X}$.)

By assumption the nilmanifold $\tilde{X} = \overline{\{\tilde{b}^n\Gamma, n \in \mathbb{N}\}}$ is connected. Since the nilrotation $\tilde{b}$ acts ergodically on the connected nilmanifold $\tilde{X}$, part of the hypothesis of Lemma 6.7 is satisfied.

Next we claim that Lemma 6.7 can be applied for the restriction of the function $\tilde{F}$ to the nilmanifold $\tilde{X}$, namely we claim that

$$\lim_{N-M\to\infty} \max_{\tilde{x}\in\tilde{X}} \left| \mathbb{E}_{M\leq n\leq N} \tilde{F}(\tilde{b}^{p(n)}\tilde{x}) - \int \tilde{F}\, dm_{\tilde{X}} \right| = 0. \tag{80}$$

Suppose first that α is rational. Since $\tilde{X}$ is a connected nilmanifold and $\tilde{\Gamma} \in \tilde{X}$, we have that $\tilde{X} \subset \{\mathbb{Z}\} \times X$. Hence, the restriction of $\tilde{F}$ onto $\tilde{X}$ is given by $\tilde{F}(\mathbb{Z}, g\Gamma) = F(g\Gamma)$ (for

[11] For example, if $T \colon \mathbb{T}^2 \to \mathbb{T}^2$ is defined by $T(x, y) = (x + \alpha, y + 2x + \alpha)$ and $k = (k_1, k_2)$, then $k \cdot T^n(x, y) = k_2\alpha n^2 + (k_1\alpha + 2k_2x)n$. Therefore, the corresponding leading term is either $k_2\alpha n^2$ (if $k_2 \neq 0$), or $k_1\alpha n$ (if $k_2 = 0$).

those $(\mathbb{Z}, g\Gamma)$ that belong to $\tilde{X}$), and as a result is continuous. In this case, (80) is a direct consequence of Lemma 6.7. Therefore, it remains to verify (80) when α is irrational. Notice first that the set of discontinuities of $\tilde{F}$ on $\tilde{X}$ is a subset of the nilmanifold $\{\mathbb{Z}\} \times X$. Near a point $(\mathbb{Z}, g\Gamma)$ of $\{\mathbb{Z}\} \times X$ the function $\tilde{F}$ comes close to the value $F(g\Gamma)$ or the value $F(b^{-1}g\Gamma)$. For $\delta > 0$ (and smaller than $1/2$) let $\tilde{X}_\delta = I_\delta \times X$ where $I_\delta = \big\{t\mathbb{Z}\colon \|t\| \geq \delta\big\}$. There exist functions $\tilde{F}_\delta \in C(\tilde{X})$ that agree with $\tilde{F}$ on $\tilde{X}_\delta$ and have sup-norm bounded by $2\,\|F\|_\infty$. Since α is irrational, the sequence $(p(n)\alpha\mathbb{Z})_{n\in\mathbb{N}}$ (which happens to be the first coordinate of $(\tilde{b}^{p(n)})_{n\in\mathbb{N}}$) is well distributed in $\mathbb{T}$, and as a result

$$\lim_{N-M\to\infty} \max_{x\in[0,1]} |\mathbb{E}_{M\leq n\leq N}\mathbf{1}_{[\delta,1-\delta]}(\{x + p(n)\alpha\}) - (1-2\delta)| = 0.$$

It follows that

$$\limsup_{N-M\to\infty} \max_{\tilde{x}\in\tilde{X}} \mathbb{E}_{M\leq n\leq N}|\tilde{F}(\tilde{b}^{p(n)}\tilde{x}) - \tilde{F}_\delta(\tilde{b}^{p(n)}\tilde{x})| \leq 3\,\|F\|_\infty\,\delta. \tag{81}$$

By Lemma 6.7, equation (80) holds if one uses the functions $\tilde{F}_\delta$ in place of the function $\tilde{F}$. Using this and (81), we deduce (80). This proves our claim.

Next notice that (80) gives that

$$\lim_{m\to\infty} \mathbb{E}_{n\in\Phi_m}\tilde{F}(\tilde{b}^{p(n)+m}\tilde{\Gamma}) = \int \tilde{F}\ dm_{\tilde{X}}. \tag{82}$$

Furthermore, since the nilrotation $\tilde{b}$ acts ergodically in $\tilde{X}$, the last integral is equal to

$$\lim_{N\to\infty} \mathbb{E}_{1\leq n\leq N}\tilde{F}(\tilde{b}^n\tilde{\Gamma}). \tag{83}$$

(If $\tilde{F}$ is not continuous we argue as before to get this.) Using (79) we see that

$$\tilde{F}(\tilde{b}^{p(n)+m}\tilde{\Gamma}) = F\big(b^{[p(n)\alpha+m\alpha]}\Gamma\big), \quad \text{and} \quad \tilde{F}(\tilde{b}^n\tilde{\Gamma}) = F(b^{[n\alpha]}\Gamma). \tag{84}$$

Combining (82), (83), and (84), we get (78). This completes the proof. $\square$

The next lemma will be used later to verify that certain connectedness assumptions (needed to apply Lemma 6.8) are satisfied.

Lemma 6.9. *Let $X = G/\Gamma$ be a connected nilmanifold and $b \in G$ be an ergodic nilrotation.*

Then there exists a connected sub-nilmanifold Z of X^ℓ such that for a.e. $g \in G$ the element $\tilde{b}_g = (g^{-1}bg, g^{-1}b^2g, \ldots, g^{-1}b^\ell g)$ acts ergodically on Z.

Remark. The independence of Z of the generic $g \in G$ will not be used, only that Z is connected will be used.

Proof. This is an immediate consequence of a limit formula that appears in Theorem 2.2 of [37] (the details of the deduction appear in Corollary 2.10 of [13]). $\square$

As mentioned before, our plan is to establish a positiveness property by comparing certain multiple ergodic averages to some simpler ones involving the sequence $[n\alpha]$. The next lemma establishes the positiveness property needed for the latter averages.

Lemma 6.10. *Let $(X, \mathcal{X}, \mu, T)$ be a system, $\ell \in \mathbb{N}$, and α be a non-zero real number.*

Then for every $f \in L^\infty(\mu)$ positive and not almost everywhere zero we have

$$\liminf_{N\to\infty} \mathbb{E}_{1\leq n\leq N} \int f \cdot T^{[n\alpha]}f \cdot \ldots \cdot T^{\ell[n\alpha]}f\ d\mu > 0. \tag{85}$$

Proof. We follow closely an argument used in [6] (Theorem 2.3).

For a general sequence $(a(n))_{n\in\mathbb{N}}$ of non-negative numbers, and $m \in \mathbb{N}$, one has

$$\liminf_{N\to\infty} \mathbb{E}_{1\le n\le N} a(n) \ge \frac{1}{m} \liminf_{N\to\infty} \mathbb{E}_{1\le n\le N} a(mn).$$

Using this for $m = 1, \dots, N_0$, where N_0 is an integer that will be chosen later, and averaging over m we get

$$\liminf_{N\to\infty} \mathbb{E}_{1\le n\le N} a(n) \ge \mathbb{E}_{1\le m\le N_0} \Big(\frac{1}{m} \liminf_{N\to\infty} \mathbb{E}_{1\le n\le N} a(mn)\Big).$$

The last expression is greater or equal than

$$\frac{1}{N_0} \liminf_{N\to\infty} \mathbb{E}_{1\le n\le N} (\mathbb{E}_{1\le m\le N_0} a(mn)).$$

As a result, if there exists $N_0 \in \mathbb{N}$ such that

$$\mathbb{E}_{1\le m\le N_0} a(mn) > c > 0$$

for a set S of $n \in \mathbb{N}$ of positive lower density, then

$$\liminf_{N\to\infty} \mathbb{E}_{1\le n\le N} a(n) > 0.$$

We shall use this for

$$a(n) = \int f \cdot T^{[n\alpha]} f \cdot \ldots \cdot T^{\ell[n\alpha]} f \; d\mu.$$

to prove (85).

We choose N_0 as follows: By the uniform multiple recurrence property ([8]), there exists $N_0 \in \mathbb{N}$, and positive constant c (depending only on $\int f d\mu$ and ℓ), such that for every $r \in \mathbb{N}$ one has

$$\mathbb{E}_{1\le m\le N_0} \int f \cdot T^{rm} f \cdot \ldots \cdot T^{\ell rm} f \; d\mu \ge c. \tag{86}$$

We choose $S = \big\{n\colon \{n\alpha\} < 1/N_0\big\}$, which as is well known has positive density. Since $[m\beta] = m[\beta]$ whenever $\{\beta\} < 1/m$, we have

$$a(mn) = \int f \cdot T^{m[n\alpha]} f \cdot \ldots \cdot T^{\ell m[n\alpha]} f \; d\mu$$

for every $n \in S$ and $m = 1, \dots, N_0$. As a result, for every $n \in S$ we have

$$\mathbb{E}_{1\le m\le N_0} a(mn) = \mathbb{E}_{1\le m\le N_0} \int f \cdot T^{m[n\alpha]} f \cdot \ldots \cdot T^{\ell m[n\alpha]} f \; d\mu \ge c$$

where the last estimate follows form (86) applied for $r = [n\alpha]$. As explained before this implies (85) and completes the proof. $\square$

We now have all the ingredients needed to prove Proposition 6.5.

Proof of Proposition 6.5. If $\alpha = 0$, or p is constant, we have $1 \prec a(t) \prec t$. Since $a \in \mathcal{H}$, it follows that for every large m the sets $\{n \in \mathbb{N}\colon [a(n)] = m\}$ are intervals with length increasing to infinity as $m \to \infty$. Using this, the result follows from Furstenberg's multiple recurrence theorem ([18]).

Suppose now that $\alpha \neq 0$ and p is non-constant. We start with some reductions. We can assume that $p(0) = 0$. Furthermore, using an ergodic decomposition argument we can assume that the system is ergodic. By Theorem 2.4 we can reduce matters to showing (70) in the case where the system is an inverse limit of nilsystems. Lastly, an argument completely analogous

to the one used in the proof of Lemma 3.2 in [20] shows that the positiveness property (70) is preserved by inverse limits. Hence, we can assume that the system is a nilsystem.

Suppose now that $X = G/\Gamma$ is a nilmanifold, $b \in G$, and $F \in L^\infty(m_X)$ is non-negative and not almost everywhere zero. Our goal is to show that there exists a sequence of intervals $(I_m)_{m\in\mathbb{N}}$ with length converging to infinity such that

$$\lim_{m\to\infty} \mathbb{E}_{n\in I_m} V([a(n)]) > 0 \tag{87}$$

where

$$V(n) = \int F(x)\cdot F(b^n x)\cdot \ldots \cdot F(b^{\ell n}x)\; dm_X. \tag{88}$$

The choice of the intervals $(I_m)_{m\in\mathbb{N}}$ will be made so that the values of the sequence $[a(n)]$ for $n \in I_m$ take a convenient form. More precisely, since $1 \prec e(t) \prec t$ and $e \in \mathcal{H}$, for every $r \in \mathbb{N}$ (to be chosen later), there exists a sequence of intervals $(J_{r,m})_{m\in\mathbb{N}}$ with $|J_{r,m}| \to \infty$, such that

$$mr\alpha \le e(n) \text{ for } n \in J_{r,m} \quad \text{and} \quad \sup_{n\in J_{r,m}} (e(n) - mr\alpha) \to 0.$$

For $\delta > 0$ (to be chosen later) define

$$S_{r,\delta} = \{n \in \mathbb{N}\colon \{p(n)r\alpha\} \le 1-\delta\}.$$

Then for every large m, if $n \in J_{r,m} \cap S_{r,\delta}$ we have

$$[a(n)] = [p(n)\alpha + mr\alpha].$$

It follows that in order to verify (87) it suffices to show that for some $r \in \mathbb{N}$ and $\delta > 0$ we have

$$\limsup_{m\to\infty} \mathbb{E}_{n\in J_{r,m}} (V([p(n)\alpha + mr\alpha]) \cdot \mathbf{1}_{S_{r,\delta}}(n)) > 0. \tag{89}$$

In fact it suffices to show that there exists $r \in \mathbb{N}$ such that

$$\limsup_{m\to\infty} \mathbb{E}_{n\in J_{r,m}} V([p(n)\alpha + mr\alpha]) > 0, \tag{90}$$

the reason being that for every $r \in \mathbb{N}$, $\alpha \in \mathbb{R}$, and δ small, we have

$$\limsup_{m\to\infty} \mathbb{E}_{n\in J_{r,m}} (1 - \mathbf{1}_{S_{r,\delta}}(n)) \le \delta,$$ [12]

and as a result (90) readily implies (89), again if δ is small. Therefore, we can concentrate our attention in proving (90).

We shall show that there exists an $r \in \mathbb{N}$ such that

$$\lim_{m\to\infty} \mathbb{E}_{n\in\Phi_m} V\big([p(rn)\alpha + mr\alpha]\big) = \lim_{N\to\infty} \mathbb{E}_{1\le n\le N} V([nr\alpha]) \tag{91}$$

for every Følner sequence $(\Phi_m)_{m\in\mathbb{N}}$ in $\mathbb{Z}$. Having shown this, we can finish the proof by using the definition of $V(n)$ and Lemma 6.10. We deduce that the limits in (91) are positive, and as a result (90) holds.

We proceed now to find this $r \in \mathbb{N}$. Using a standard lifting argument we can also assume that the group G is connected and simply connected.[13] A standard approximation argument shows that it suffices to verify (91) whenever the function F in (88) is continuous. Our plan is

[12] If α is irrational the estimate holds with equality (this follows from equidistribution). If α is rational the estimate holds trivially for δ small.

[13] We can assume that G/G_0 is finitely generated. In this case, one can show ([28]) that there exists a nilmanifold $\hat{X} = \hat{G}/\hat{\Gamma}$, where $\hat{G}$ is a connected and simply-connected Lie group, such that for every $F \in C(X)$, $b \in G$, and $x \in X$, there exists $\hat{F} \in C(\hat{X})$, $\hat{b} \in \hat{G}$, and $\hat{x} \in \hat{X}$, such that $F(b^n x) = \hat{F}(\hat{b}^n \hat{x})$ for every $n \in \mathbb{N}$.

to use Lemma 6.8 to establish a stronger pointwise result, namely, there exists an $r \in \mathbb{N}$ such that for a.e. $x \in X$ we have for every Følner sequence $(\Phi_m)_{m\in\mathbb{N}}$ in $\mathbb{Z}$ that

(92)
$$\lim_{m\to\infty} \mathbb{E}_{n\in\Phi_m} F(b^{[p(rn)\alpha+mr\alpha]}x)\cdot\ldots\cdot F(b^{\ell[p(rn)\alpha+mr\alpha]}x) = \lim_{m\to\infty} \mathbb{E}_{1\le n\le N} F(b^{[nr\alpha]}x)\cdot\ldots\cdot F(b^{\ell[nr\alpha]}x).$$

(Our argument will show that both limits exist.) Trivially, this pointwise result implies (91). Therefore, we are left with establishing (92).

We do some preparation in order to use Lemma 6.8. For every $n \in \mathbb{N}$ and element $x = g\Gamma$ of X we have

$$F(b^n x)\cdot\ldots\cdot F(b^{\ell n}x) = \tilde{F}(\tilde{b}_g^n \tilde{\Gamma}), \tag{93}$$

where $\tilde{F}(x_1,\ldots,x_l) = F(gx_1)\cdot\ldots\cdot F(gx_\ell)$ $(\in C(X^\ell))$, $\tilde{\Gamma} = \Gamma^\ell$, and

$$\tilde{b}_g = (g^{-1}bg, g^{-1}b^2g, \ldots, g^{-1}b^\ell g) \in G^\ell.$$

Next we show that there exists an $r \in \mathbb{N}$ such that for a.e. $g \in G$ the sub-nilmanifold

$$\overline{\{(nr\alpha, \tilde{b}_g^{nr\alpha}\tilde{\Gamma}), n \in \mathbb{N}\}} \tag{94}$$

of $\mathbb{T}\times X^\ell$ is connected. Indeed, as mentioned in Section 3.2, there exists r_0 such that the nilmanifold $\overline{\{(nr_0\alpha\mathbb{Z}, b^{nr_0\alpha}\Gamma), n \in \mathbb{N}\}}$ is connected. It follows form Lemma 6.9 (applied to the nilmanifold $\mathbb{T}\times X$ and the element $(r_0\alpha, b^{r_0\alpha}) \in \mathbb{T}\times G$) that for almost every $g \in G$ the sub-nilmanifold

$$\overline{\{(nr_0\alpha\mathbb{Z},\ldots,\ell nr_0\alpha\mathbb{Z}, \tilde{b}_g^{nr_0\alpha}\tilde{\Gamma}), n \in \mathbb{N}\}}$$

of $\mathbb{T}^\ell\times X^\ell$ is connected. Projecting in the appropriate coordinates we get that the nilmanifold in (94) is also connected.

We are now in a position where we can apply Lemma 6.8 for the nilmanifold X^ℓ, the function $\tilde{F} \in C(X^\ell)$, the integer r_0 we just found, the element $r_0\alpha \in \mathbb{R}$ in place of α, the polynomial $q \in \mathbb{Z}[t]$ defined by $q(t) = p(r_0t)/r_0$ (remember $p(0) = 0$), and the elements $\tilde{b}_g \in G^\ell$ (for those $g \in G$ for which the set (94) is connected for $r = r_0$). We deduce that for almost every $g \in G$, and for every Følner sequence $(\Phi_m)_{m\in\mathbb{N}}$ in $\mathbb{Z}$, we have that

$$\lim_{m\to\infty} \mathbb{E}_{n\in\Phi_m} \tilde{F}(\tilde{b}_g^{[p(r_0n)\alpha+mr_0\alpha]}\tilde{\Gamma}) = \lim_{N\to\infty} \mathbb{E}_{1\le n\le N} \tilde{F}(\tilde{b}_g^{[nr_0\alpha]}\tilde{\Gamma}).$$

Using (93), this gives (92), which in turn gives (91). This completes the proof. □

References

[1] T. Austin. On the norm convergence of non-conventional averages. *Ergodic Theory Dynam. Systems* **30** (2010), 321–338.

[2] B. Berend. Joint ergodicity and mixing. *J. Analyse Math.* **4**5 (1985), 255–284.

[3] V. Bergelson. Weakly mixing PET. *Ergodic Theory Dynam. Systems* **7** (1987), no. 3, 337–349.

[4] V. Bergelson. Ergodic Ramsey Theory. *Logic and Combinatorics* (editted by S. Simpson). *Contemporary Mathematics* **6**5 (1987), 63–87.

[5] V. Bergelson, M. Boshernitzan, J. Bourgain. Some results on nonlinear recurrence. *J. Anal. Math.* **62** (1994), 29–46.

[6] V. Bergelson, I. Håland. Sets of recurrence and generalized polynomials. Convergence in ergodic theory and probability (Columbus, OH, 1993), *Ohio State Univ. Math. Res. Inst. Publ.*, **5**, de Gruyter, Berlin, (1996), 91-110.

[7] V. Bergelson, I. Håland-Knutson. Weak mixing implies mixing of higher orders along tempered functions. *Ergodic Theory Dynam. Systems* **29** (2009), no. 5, 1375–1416

[8] V. Bergelson, B. Host, R. McCutcheon, F. Parreau. Aspects of uniformity in recurrence. *Colloq. Math.* **84/85** (2000), no. 2, 549–576.

[9] V. Bergelson, A. Leibman. Polynomial extensions of van der Waerden's and Szemerédi's theorems. *J. Amer. Math. Soc.* **9** (1996), 725–753.

[10] V. Bergelson, A. Leibman, E. Lesigne. Intersective polynomials and polynomial Szemeredi theorem. *Adv. Math.* **219** (2008), no. 1, 369–388.

[11] M. Boshernitzan. Uniform distribution and Hardy fields. *J. Anal. Math.* **62** (1994), 225–240.

[12] M. Boshernitzan, G. Kolesnik, A. Quas, M. Wierdl. Ergodic averaging sequences. *J. Anal. Math.* **95** (2005), 63–103.

[13] N. Frantzikinakis. Multiple ergodic averages for three polynomials and applications. *Trans. Amer. Math. Soc.* **360** (2008), no. 10, 5435-5475.

[14] N. Frantzikinakis. Equidistribution of sparse sequences on nilmanifolds. *J. Analyse Math.* **109** (2009), 1–43.

[15] N. Frantzikinakis, B. Host, B. Kra. Multiple recurrence and convergence for sets related to the primes. *J. Reine Angew. Math.* **611** (2007), 131–144.

[16] N. Frantzikinakis, B. Kra. Ergodic averages for independent polynomials and applications. *J. London Math. Soc.* **74** (2006), no. 1, 131–142.

[17] N. Frantzikinakis, M. Wierdl. A Hardy field extension of Szemerédi's theorem. *Advances in Math.* **222** (2009), 1–43.

[18] H. Furstenberg. Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions. *J. Analyse Math.* **71** (1977), 204–256.

[19] H. Furstenberg. Recurrence in ergodic theory and combinatorial number theory. *Princeton University Press*, Princeton, (1981).

[20] H. Furstenberg, Y. Katznelson. An ergodic Szemerédi theorem for commuting transformations. *J. Analyse Math.* **34** (1979), 275–291.

[21] B. Green, T. Tao. The primes contain arbitrarily long arithmetic progressions. *Annals. Math.* **167** (2008), 481–547.

[22] B. Green, T. Tao. The quantitative behaviour of polynomial orbits on nilmanifolds. Preprint 2007. Available at arXiv:0709.3562

[23] G. Hardy. Properties of logarithmico-exponential functions. *Proc. London Math. Soc.* (2) **10** (1912), 54–90.

[24] G. Hardy. Orders of Infinity. *Cambridge Tracts in Math. and Math. Phys.*, 12 (2nd edition), Cambridge, (1924).

[25] B. Host. Ergodic seminorms for commuting transformations and applications. *Studia Math.* **195** (1) (2009), 31–49.

[26] B. Host, B. Kra. Nonconventional ergodic averages and nilmanifolds. *Annals Math.* **161** (2005), 397–488.

[27] B. Host, B. Kra. Convergence of polynomial ergodic averages. *Isr. J. Math.* **149** (2005), 1–19.

[28] A. Leibman. Pointwise convergence of ergodic averages for polynomial sequences of rotations of a nilmanifold. *Ergodic Theory Dynam. Systems* **25** (2005), no. 1, 201–213.

[29] A. Leibman. Convergence of multiple ergodic averages along polynomials of several variables. *Isr. J. Math.* **146** (2005), 303–316.

[30] A. Leibman. Orbit of the diagonal in the power of a nilmanifold. *Trans. Amer. Math. Soc.* **362** (2010), 1619–1658.

[31] K. Petersen. Ergodic theory. *Cambridge Studies in Advanced Mathematics*, **2**. Cambridge University Press, Cambridge, (1989).

[32] E. Szemerédi. On sets of integers containing no k elements in arithmetic progression. *Acta Arith.* **27** (1975), 299–345.

[33] T. Tao. Norm convergence of multiple ergodic averages for commuting transformations. *Ergodic Theory Dynam. Systems* **28** (2008), no. 2, 657–688.

[34] T. Tao, T. Ziegler. The primes contain arbitrarily long polynomial progressions. *Acta Math.* **201** (2008), 213–305.

[35] H. Towsner. Convergence of diagonal ergodic averages. *Ergodic Theory Dynam. Systems* **29** (2009), 1309–1326.

[36] P. Walters. An introduction to ergodic theory. *Graduate Texts in Mathematics*, **79**, Springer-Verlag, New York-Berlin, (1982).

[37] T. Ziegler. A non-conventional ergodic theorem for a nilsystem, *Ergodic Th. Dynam. Systems* **25** (2005), no. 4, 1357–1370.

[38] T. Ziegler. Universal characteristic factors and Furstenberg averages. *J. Amer. Math. Soc.* **20** (2007), 53–97.

(Nikos Frantzikinakis) Department of Mathematics, University of Crete, Knossos Avenue, Heraklion, 71409, GREECE

E-mail address: frantzikinakis@gmail.com